

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2005, 01:08:33 ; Search time 759 Seconds  
(without alignments)  
8992.812 Million cell updates/sec

Title: US-10-829-118-1  
Perfect score: 980  
Sequence: 1 ctccgcggcgacacgcctg.....ctgcaccacggaatgctga 980

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 8443130 seqs, 3482420727 residues

Total number of hits satisfying chosen parameters: 16986260

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*
- 20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*
- 21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*
- 22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*
- 24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	980	100.0	980	9	US-09-891-711-1
2	980	100.0	980	19	US-10-829-118-1
3	980	100.0	2469	9	US-09-891-711-2
4	980	100.0	2469	19	US-10-829-118-2
c 5	124.4	12.7	4338	9	US-09-891-711-3
c 6	124.4	12.7	4338	17	US-10-172-118-953
c 7	124.4	12.7	4338	18	US-10-342-887-953

c 8	124.4	12.7	4338	19	US-10-829-118-3	Sequence 3, Appli
c 9	124.4	12.7	4338	21	US-10-956-157-1116	Sequence 1116, Ap
c 10	56.4	5.8	1523	19	US-10-767-701-9786	Sequence 9786, Ap
c 11	53	5.4	658	19	US-10-767-701-4204	Sequence 4204, Ap
c 12	52.6	5.4	4667	20	US-10-723-860-5759	Sequence 5759, Ap
c 13	51.6	5.3	598	20	US-10-425-115-2989	Sequence 2989, Ap
c 14	51.6	5.3	763	19	US-10-437-963-49370	Sequence 49370, A
c 15	51.4	5.2	606	18	US-10-424-599-95310	Sequence 95310, A
c 16	50.8	5.2	2307	10	US-09-893-519A-87	Sequence 87, Appli
c 17	49.8	5.1	849	20	US-10-363-345A-16551	Sequence 16551, A
c 18	49.8	5.1	849	20	US-10-363-345A-16552	Sequence 16552, A
c 19	49.8	5.1	849	21	US-10-363-483A-16551	Sequence 16551, A
c 20	49.8	5.1	849	21	US-10-363-483A-16552	Sequence 16552, A
c 21	48.8	5.0	771	14	US-10-010-920-4	Sequence 4, Appli
c 22	48.8	5.0	771	14	US-10-010-920-5	Sequence 5, Appli
c 23	48.8	5.0	771	14	US-10-008-721-4	Sequence 4, Appli
c 24	48.8	5.0	771	14	US-10-008-721-5	Sequence 5, Appli
c 25	48.8	5.0	955	14	US-10-010-920-3	Sequence 3, Appli
c 26	48.8	5.0	955	14	US-10-008-721-3	Sequence 3, Appli
c 27	48.8	5.0	40668	19	US-10-450-826-57	Sequence 57, Appli
c 28	48.6	5.0	809	18	US-10-424-599-26988	Sequence 26988, A
c 29	48.4	4.9	4225	16	US-10-141-541-3	Sequence 3, Appli
c 30	48.2	4.9	1187	16	US-10-029-386-25125	Sequence 25125, A
c 31	48	4.9	1092	19	US-10-767-701-9739	Sequence 9739, Ap
c 32	48	4.9	1215	19	US-10-767-701-8767	Sequence 8767, Ap
c 33	48	4.9	38847	13	US-10-087-192-868	Sequence 868, App
c 34	48	4.9	150275	22	US-10-981-277-55	Sequence 55, Appli
c 35	47.8	4.9	951	19	US-10-437-963-63083	Sequence 63083, A
c 36	47.6	4.9	1041	19	US-10-437-963-8597	Sequence 8597, Ap
c 37	47.4	4.8	154746	10	US-09-827-688-8	Sequence 8, Appli
c 38	47.4	4.8	154746	10	US-09-827-688-8	Sequence 8, Appli
c 39	47.2	4.8	825	19	US-10-437-963-60378	Sequence 60378, A
c 40	47	4.8	876	20	US-10-425-115-55664	Sequence 55664, A
c 41	47	4.8	876	20	US-10-425-115-55664	Sequence 55664, A
c 42	46.8	4.8	1211	19	US-10-437-963-27585	Sequence 27585, A
c 43	46.8	4.8	2571	14	US-10-277-032-1	Sequence 1, Appli
c 44	46.8	4.8	2571	18	US-10-681-223-1	Sequence 1, Appli
c 45	46.8	4.8	2856	16	US-10-305-810-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-891-711-1  
; Sequence 1, Application US/09891711  
; Patent No. US20020082404A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Promoter Sequences  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/09/891,711  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 980  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-891-711-1

Query Match 100.0%; Score 980; DB 9; Length 980;

Best Local Similarity 100.0%; Pred. No. 1.7e-288; Indels 0; Gaps 0;

Matches 980; Conservative 0; Mismatches 0;

Qy	1	CTCCGCGGCGAACACGCCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCCGAG	60
Db	1	CTCCGCGGCGAACACGCCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCCGAG	60
Qy	61	ACTGGGCGGCGGGCGGCGAGGCCACAGCTGGGAGCCTCAGCTCCGCCGACCCAGCGT	120
Db	61	ACTGGGCGGCGGGCGGCGAGGCCACAGCTGGGAGCCTCAGCTCCGCCGACCCAGCGT	120

```
QY 121 GCCCTGTCTGTCCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGACGCGCTGGCAGCGG 180
DB 121 GCCCTGTCTGTCCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGACGCGCTGGCAGCGG 180
QY 181 GACACCGCGGAGGATGGACGAAGTGTCTCGACATTTGGCGGGCGGGCGGGTGG 240
DB 181 GACACCGCGGAGGATGGACGAAGTGTCTCGACATTTGGCGGGCGGGCGGGTGG 240
QY 241 CAGGGTGAAGCGGAGGGCGTGGCCAGCAGCTGCCAGGCGGAGAACCGCTGGGGG 300
DB 241 CAGGGTGAAGCGGAGGGCGTGGCCAGCAGCTGCCAGGCGGAGAACCGCTGGGGG 300
QY 301 AACCTTTGGTCCGCTCTGCGGTCTAGATATCCCGAAAGAGACACGGGCGCGAAA 360
DB 301 AACCTTTGGTCCGCTCTGCGGTCTAGATATCCCGAAAGAGACACGGGCGCGAAA 360
QY 361 CGGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGGG 420
DB 361 CGGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGGG 420
QY 421 GCCTGTCTCATGACTCGTGGAGGCTCGCTTCCCGCGGACCCCTTCTTCCAGGGTGC 480
DB 421 GCCTGTCTCATGACTCGTGGAGGCTCGCTTCCCGCGGACCCCTTCTTCCAGGGTGC 480
QY 481 CAGCTCCAGGACACCGGCGGCTCGGACACCCACCCCGCGGCGACCTGCGCTGGGTGC 540
DB 481 CAGCTCCAGGACACCGGCGGCTCGGACACCCACCCCGCGGCGACCTGCGCTGGGTGC 540
QY 541 CCTTTAACCCCGGCGGTAGCTCGTTTAAAGATGGCGAAGTGTCCGGTCCCGAAACACGCGAAA 600
DB 541 CCTTTAACCCCGGCGGTAGCTCGTTTAAAGATGGCGAAGTGTCCGGTCCCGAAACACGCGAAA 600
QY 601 CCCCAATCCCGCTGCCGAGCTCTCGACCCCGCGGCGGACACGAGACTGGGCC 660
DB 601 CCCCAATCCCGCTGCCGAGCTCTCGACCCCGCGGCGGACACGAGACTGGGCC 660
QY 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGCAGAACCGAGGACCTTTGTAAAC 720
DB 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGCAGAACCGAGGACCTTTGTAAAC 720
QY 721 GCCAGTGTGCTCTTTTGAAGAAACAGAAATAAATGTGTAAAGCTCTGAAAGCT 780
DB 721 GCCAGTGTGCTCTTTTGAAGAAACAGAAATAAATGTGTAAAGCTCTGAAAGCT 780
QY 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTCAATGTAAA 840
DB 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTCAATGTAAA 840
QY 841 TGACACGCGGAGCTCCGCTACAGGCGGATCAGACAGTTTGTGTCTTCTGGAACCTTGACA 900
DB 841 TGACACGCGGAGCTCCGCTACAGGCGGATCAGACAGTTTGTGTCTTCTGGAACCTTGACA 900
QY 961 CTGACACGCGTAAGTCTGA 980
DB 961 CTGACACGCGTAAGTCTGA 980
```

## RESULT 2

US-10-829-118-1

```
/ Sequence 1, Application US/10829118
/ Publication No. US20040191759A1
/ GENERAL INFORMATION:
/ APPLICANT: Pharmacia AB
/ TITLE OF INVENTION: Promoter Sequences
/ FILE REFERENCE: 00130
/ CURRENT APPLICATION NUMBER: US/10/829,118
/ CURRENT FILING DATE: 2004-04-20
/ PRIOR APPLICATION NUMBER: US/09/891,711
/ PRIOR FILING DATE: 2001-06-26
```

```
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 980
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-829-118-1
```

```
Query Match 100.0%; Score 980; DB 19; Length 980;
Best Local Similarity 100.0%; Pred. No. 1.7e-288;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CTCCGCGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
DB 1 CTCCGCGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
QY 61 ACTGGGCGGGCGGCGAGGAGCCACAGCTGGGAGCCTCAGCTCCGCCACCCAGCGT 120
DB 61 ACTGGGCGGGCGGCGAGGAGCCACAGCTGGGAGCCTCAGCTCCGCCACCCAGCGT 120
QY 121 GCCCTGTCTGTCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGACGCGCTGGGAGCGG 180
DB 121 GCCCTGTCTGTCTCCGCGCTCCCGGGCTTGGTGGCGCTCTGACGCGCTGGGAGCGG 180
QY 181 GACACGCGGGAGGATGGACGAAGTGTCTCGACATTTTGGCGGGCGGGCGGGTGG 240
DB 181 GACACGCGGGAGGATGGACGAAGTGTCTCGACATTTTGGCGGGCGGGCGGGTGG 240
QY 241 CAGGGTGAAGCGGAGGGCGTGGCCAGCAGCTGCCAGGCGGAGAACCGCTGGGGG 300
DB 241 CAGGGTGAAGCGGAGGGCGTGGCCAGCAGCTGCCAGGCGGAGAACCGCTGGGGG 300
QY 301 AACCTTTGGTCCGCTCTGCGGTCTAGGATCCCGAAAGAGACACGGGCGCGAAA 360
DB 301 AACCTTTGGTCCGCTCTGCGGTCTAGGATCCCGAAAGAGACACGGGCGCGAAA 360
QY 361 CGGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGGG 420
DB 361 CGGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCGCGGG 420
QY 421 GCCTGTCTCATGACTCGTGGAGGCTCGCTTCCCGCGGACCCCTTCTTCCAGGGTGC 480
DB 421 GCCTGTCTCATGACTCGTGGAGGCTCGCTTCCCGCGGACCCCTTCTTCCAGGGTGC 480
QY 481 CAGCTCCAGGACACCGGCGGCTCGGACACCCACCCCGCGGCGGACCTGCGCTGGGTGC 540
DB 481 CAGCTCCAGGACACCGGCGGCTCGGACACCCACCCCGCGGCGGACCTGCGCTGGGTGC 540
QY 541 CCTTTAACCCCGGCGGTAGCTCGTTTAAAGATGGCGAAGTGTCCGGTCCCGAAACACGCGAAA 600
DB 541 CCTTTAACCCCGGCGGTAGCTCGTTTAAAGATGGCGAAGTGTCCGGTCCCGAAACACGCGAAA 600
QY 601 CCCCAATCCCGCTGCCGAGCTCTCGACCCCGCGGCGGACACGAGACTGGGCC 660
DB 601 CCCCAATCCCGCTGCCGAGCTCTCGACCCCGCGGCGGACACGAGACTGGGCC 660
QY 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGCAGAACCGAGGACCTTTGTAAAC 720
DB 661 TCCGACGCGAGCGCTGCGGAGCACCGGTGCGTGCAGAACCGAGGACCTTTGTAAAC 720
QY 721 GCCAGTGTGCTCTTTTGAAGAAACAGAAATAAATGTGTAAAGCTCTGAAAGCT 780
DB 721 GCCAGTGTGCTCTTTTGAAGAAACAGAAATAAATGTGTAAAGCTCTGAAAGCT 780
QY 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTCAATGTAAA 840
DB 781 TGCGCCCTAAAGATGTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTCAATGTAAA 840
QY 841 TGACACGCGGAGCTCCGCTACAGGCGGATCAGACAGTTTGTGTCTTCTGGAACCTTGACA 900
DB 841 TGACACGCGGAGCTCCGCTACAGGCGGATCAGACAGTTTGTGTCTTCTGGAACCTTGACA 900
QY 901 TGGTGGGCGGTGATGCTGATCAGGCGGATCAGACAGTTTGTGTCTTCTGGAACCTTGACA 960
```

Db 901 TGGTGGCGGTGATCGCTACAGCCGATCAGACAGTTTGTGCTCTCTGGAACCTTGACA 960  
Qy 961 CTGCACACCGTAAATGCTGA 980  
Db 961 CTGCACACCGTAAATGCTGA 980  
RESULT 3  
US-09-891-711-2  
; Sequence 2, Application US/09891711  
; Patent No. US20020082404A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/09/891,711  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2469  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-891-711-2

Query Match 100.0%; Score 980; DB 9; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 2.3e-288;  
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CTCGCGGGGAAACACGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
Db 353 CTCGCGGGGAAACACGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 412  
Qy 61 ACTGGGGCGGGCCAGCGAGGCCACACAGCTGGAGCCTCAGCTCGCGGACCCAGCGT 120  
Db 413 ACTGGGGCGGGCCAGCGAGGCCACACAGCTGGAGCCTCAGCTCGCGGACCCAGCGT 472  
Qy 121 GCCTGTCTGTCTCCGCGCTCCCGGGCTTCGCGCGCTCTGAGCGCGCTGCGGAGCGG 180  
Db 473 GCCTGTCTGTCTCCGCGCTCCCGGGCTTCGCGCGCTCTGAGCGCGCTGCGGAGCGG 532  
Qy 181 GACCACCGCGGAGGATGGAAGAAGTGTCTCGGACATTTTCGGCGCGCGGGCGCGGTGG 240  
Db 533 GACCACCGCGGAGGATGGAAGAAGTGTCTCGGACATTTTCGGCGCGCGGGCGCGGTGG 592  
Qy 241 CAGGGTGGAAAGCGGAGGGCGTGGCCAGCGAGCTGCCAGCGCGAGAACCGCTGGGGG 300  
Db 593 CAGGGTGGAAAGCGGAGGGCGTGGCCAGCGAGCTGCCAGCGCGAGAACCGCTGGGGG 652  
Qy 301 AACCCCTTGTCTCGCTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 360  
Db 653 AACCCCTTGTCTCGCTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 712  
Qy 361 GCGGCCAGGCTGGGCGAGGATCTAGAAAGACTTCCTTGGCGCAGGCTCCCTGCCCGCGCGG 420  
Db 713 GCGGCCAGGCTGGGCGAGGATCTAGAAAGACTTCCTTGGCGCAGGCTCCCTGCCCGCGG 772  
Qy 421 GCCTGTCTGATGGAAGTCTGTTGAGAGCTCGCTTCCGCGCGGACCTTCTCTGAGGGGTC 480  
Db 773 GCCTGTCTGATGGAAGTCTGTTGAGAGCTCGCTTCCGCGCGGACCTTCTCTGAGGGGTC 832  
Qy 481 CAGTCCAGGACCGGGGCTCGGACACCCACCCCGCGGCGACCTGCCCTGGGTGC 540  
Db 833 CAGTCCAGGACCGGGGCTCGGACACCCACCCCGCGGCGACCTGCCCTGGGTGC 592  
Qy 541 CCCTTAACCCGGGCGGTAGTCTGTTAAGATGGCGAAAGTGTCCGCTCCGGAACACGCGAAA 600  
Db 893 CCCTTAACCCGGGCGGTAGTCTGTTAAGATGGCGAAAGTGTCCGCTCCGGAACACGCGAAA 952  
Qy 601 CCCCAAATCCCGCTCCCGACCTCTGACCCCGCGGCCACCGGACGACAGACTGGGCC 660  
Db 953 CCCCAAATCCCGCTCCCGACCTCTGACCCCGCGGCCACCGGACGACAGACTGGGCC 1012

Qy 661 TCCCGACGCGCAGCGCGCTGCCGGGACACCGGTGCGTGGGAAACGGAGGACCTTTGTAAAC 720  
Db 1013 TCCCGACGCGCAGCGCGCTGCCGGGACACCGGTGCGTGGGAAACGGAGGACCTTTGTAAAC 1072  
Qy 721 GCCACGTGTTGCTCTTTTGTGAAAAAACAAGATAAATGTTAAACTGTCGTAAGAGCT 780  
Db 1073 GCCACGTGTTGCTCTTTTGTGAAAAAACAAGATAAATGTTAAACTGTCGTAAGAGCT 1132  
Qy 781 TGCCGCGCTAAAGATGCTCTGGGTGACTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAAA 840  
Db 1133 TGCCGCGCTAAAGATGCTCTGGGTGACTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAAA 1192  
Qy 841 TGGACCAAGCCCGGACTCCGCTACGGCACTAGCAGGGGACTGAAAGCGCTTTTCAGGTACTGC 900  
Db 1193 TGGACCAAGCCCGGACTCCGCTACGGCACTAGCAGGGGACTGAAAGCGCTTTTCAGGTACTGC 1252  
Qy 901 TGGTGGGGCGGTGATGCGCTACAGGCCGATCAGACAGTTTGTGTCCTTCTGGAACCTTGACA 960  
Db 1253 TGGTGGGGCGGTGATGCGCTACAGGCCGATCAGACAGTTTGTGTCCTTCTGGAACCTTGACA 1312  
Qy 961 CTGCACACCGTAAATGCTGA 980  
Db 1313 CTGCACACCGTAAATGCTGA 1332

RESULT 4  
US-10-829-118-2  
; Sequence 2, Application US/10829118  
; Publication No. US20040191759A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Promoter Sequences  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/10/829,118  
; PRIOR FILING DATE: 2004-04-20  
; PRIOR APPLICATION NUMBER: US/09/891,711  
; PRIOR FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2469  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-829-118-2

Query Match 100.0%; Score 980; DB 19; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 2.3e-288;  
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CTCGCGGGGAAACACGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
Db 353 CTCGCGGGGAAACACGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 412  
Qy 61 ACTGGGGCGGGCCAGCGAGGCCACACAGCTGGAGCCTCAGCTCCGCGGACCCAGCGT 120  
Db 413 ACTGGGGCGGGCCAGCGAGGCCACACAGCTGGAGCCTCAGCTCCGCGGACCCAGCGT 472  
Qy 121 GCCTGTCTGTCTCCGCGCTCCCGGGCTTCGCGCGCTCTGGAAGCGCGTGGGCGAGCGG 180  
Db 473 GCCTGTCTGTCTCCGCGCTCCCGGGCTTCGCGCGCTCTGGAAGCGCGTGGGCGAGCGG 532  
Qy 181 GACCACCGCGGAGGATGGAAGAAGTGTCTCGGACATTTTCGGCGCGGGCGCGGTGG 240  
Db 533 GACCACCGCGGAGGATGGAAGAAGTGTCTCGGACATTTTCGGCGCGGGCGCGGTGG 592  
Qy 241 CAGGGTGGAAAGCGGAGGGCGTGGCCAGCGAGCTGCCAGCGCGGAGAACCGCTGGGGG 300  
Db 593 CAGGGTGGAAAGCGGAGGGCGTGGCCAGCGAGCTGCCAGCGCGGAGAACCGCTGGGGG 652  
Qy 301 AACCCCTTGTCTCGCTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 360  
Db 653 AACCCCTTGTCTCGCTCTGCGCGTCTAGGATCCCGAAAGGAGGACGCGGCGGAAA 712

361 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTGCGCGAGGCTCCCTGCCCCCGGG 420  
Db |||||  
713 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTGCGCGAGGCTCCCTGCCCCCGGG 772  
Qy |||||  
421 GCTGCTGTCATGGAAGCTGCGGAGAGCTGCTTCCCGCGGAGACCTTCTTCAGGGGTC 480  
Db |||||  
773 GCTGCTGTCATGGAAGCTGCGGAGAGCTGCTTCCCGCGGAGACCTTCTTCAGGGGTC 832  
Qy |||||  
481 CACGTCCAGGACCGGCGGCTCGGACACCCACCCCGCGCGGACCTGCGCTCGGGTGC 540  
Db |||||  
833 CACGTCCAGGACCGGCGGCTCGGACACCCACCCCGCGCGGACCTGCGCTCGGGTGC 892  
Qy |||||  
541 CCTTTAACCCCGGCGGCTAGCTCGTTAAGATGCGAAGTGTCCGGTCCGGAAACACGCGAAA 600  
Db |||||  
893 CCTTTAACCCCGGCGGCTAGCTCGTTAAGATGCGAAGTGTCCGGTCCGGAAACACGCGAAA 952  
Qy |||||  
601 CCCCAGATCCCGCTGCGGAGCTCTGTGACCCCGCGGCCCGACGCGGACGACACTGGGCC 660  
Db |||||  
953 CCCCAGATCCCGCTGCGGAGCTCTGTGACCCCGCGGCCCGACGCGGACGACACTGGGCC 1012  
Qy |||||  
661 TCCGACGCGGACGCGGCTGCGGGGACACCGGTGCGTGGGAAACGAGGACCTTTGTAAAC 720  
Db |||||  
1013 TCCGACGCGGACGCGGCTGCGGGGACACCGGTGCGTGGGAAACGAGGACCTTTGTAAAC 1072  
Qy |||||  
721 GCCAGCTGTTGCTCTTTTGGAAAAAACAAGATAAATGTGTTAACTGTCTGAAAAGCT 780  
Db |||||  
1073 GCCAGCTGTTGCTCTTTTGGAAAAAACAAGATAAATGTGTTAACTGTCTGAAAAGCT 1132  
Qy |||||  
781 TCGCGCTTAAAGATGCTGCGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840  
Db |||||  
1133 TCGCGCTTAAAGATGCTGCGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 1192  
Qy |||||  
841 TGGACACGCGGACTCCGCTAGCGGACTAGCGGAGCTGAAAGCGCTCTTCAGGTACTGC 900  
Db |||||  
1193 TGGACACGCGGACTCCGCTAGCGGACTAGCGGAGCTGAAAGCGCTCTTCAGGTACTGC 1252  
Qy |||||  
901 TGGTGGCGGCTGATGCGCTACAGCGGATCAGACAGTTTGTCTTCTGGAACCTTGACA 960  
Db |||||  
1253 TGGTGGCGGCTGATGCGCTACAGCGGATCAGACAGTTTGTCTTCTGGAACCTTGACA 1312  
Qy |||||  
961 CTGCACACGCTTAATCTCTGA 980  
Db |||||  
1313 CTGCACACGCTTAATCTCTGA 1332  
RESULT 5  
US-09-891-711-3/c  
; Sequence 3, Application US/09891711  
; Patent No. US20020082404A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Promoter Sequences  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/09/891,711  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 4338  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (497)..(3655)  
US-09-891-711-3  
Query Match 12.7%; Score 124.4; DB 9; Length 4338;  
Best Local Similarity 99.2%; Pred. No. 5.4e-27;  
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CTCCGCGGGAACACCGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCCGCGAG 60  
Db |||||  
RESULT 6  
US-10-829-118-1/c  
; Sequence 3, Application US/10172118  
; Publication No. US20030224374A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Chris  
; APPLICANT: Van 't Veer, Laura  
; APPLICANT: Van de Vijver, Marc  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-175-999  
; CURRENT APPLICATION NUMBER: US/10/172,118  
; CURRENT FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: 60/380,770  
; PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 2699  
; SEQ ID NO 953  
; LENGTH: 4338  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: NM\_003791  
; DATABASE ENTRY DATE: 2001-06-18  
US-10-172-118-953  
Query Match 12.7%; Score 124.4; DB 17; Length 4338;  
Best Local Similarity 99.2%; Pred. No. 5.4e-27;  
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CTCCGCGGGAACACCGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCCGCGAG 60  
Db |||||  
126 CTCCGCGGGAACACCGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCCGCGAG 67  
Qy |||||  
61 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
Db |||||  
66 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 127  
Qy |||||  
121 GCCCTG 126  
Db 6 GCCCTG 1  
RESULT 7  
US-10-342-887-953/c  
; Sequence 953, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Christopher J.  
; APPLICANT: Van 't Veer, Laura Johanna  
; APPLICANT: Van de Vijver, Marc J.  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-198-999  
; CURRENT APPLICATION NUMBER: US/10/342,887



```
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 953
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-953

Query Match      12.7%; Score 124.4; DB 18; Length 4338;
Best Local Similarity 99.2%; Pred. No. 5.4e-27;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCGCGGAACACAGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGCGGAACACAGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67

QY 61 ACTGGGCGCGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 66 ACTGGGCGGACCGGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7

QY 121 GCCCTG 126
Db 6 GCCCTG 1

*RESULT 8
US-10-829-118-3/c
; Sequence 3, Application US/10829118
; Publication No. US20040191759A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/10/829,118
; CURRENT FILING DATE: 2004-04-20
; PRIOR APPLICATION NUMBER: US/09/891,711
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (497)..(3655)
US-10-829-118-3
```

```
Query Match      12.7%; Score 124.4; DB 19; Length 4338;
Best Local Similarity 99.2%; Pred. No. 5.4e-27;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCGCGGAACACAGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGCGGAACACAGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67

QY 61 ACTGGGCGCGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 66 ACTGGGCGGACCGGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7

QY 121 GCCCTG 126
Db 6 GCCCTG 1

RESULT 9
US-10-956-157-1116/c
```

```
; Sequence 1116, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1116
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-1116

Query Match      12.7%; Score 124.4; DB 21; Length 4338;
Best Local Similarity 99.2%; Pred. No. 5.4e-27;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCGCGCGGAACACAGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGCGGAACACAGCGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67

QY 61 ACTGGGCGCGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 66 ACTGGGCGGACCGGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7

QY 121 GCCCTG 126
Db 6 GCCCTG 1

RESULT 10
US-10-767-701-9786/c
; Sequence 9786, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 9786
; LENGTH: 1523
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS34446_1
US-10-767-701-9786

Query Match      5.8%; Score 56.4; DB 19; Length 1523;
Best Local Similarity 51.6%; Pred. No. 2.2e-06;
Matches 129; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 133 CCGCGCTCCCGGGCTTGGTGGCGCTCTGGACGCCGTGGGCGAGCGGGACCAACGCGGG 192
Db 279 CGCGGCGCAACGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 220

QY 193 AGGATGGACGAAGGTCTCGCGACATTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 252
Db 219 GGACCGCGGGAGGAGCCCGCGGCTCGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 160

QY 253 GGAGGGCGCTGGCCAGCAGCTGCGGCGGCGGAGAACCGCGCTGGGGGGAACCTTGGTTC 312
Db 159 GGCGCGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 100
```



QY 182 ACCACGCGGAGGATGACAAAGTGCTCGCGACATTTTCGGCGGCGGGGCGCGGTGGC 241  
Db 376 GGGGCGGCCCCCGCGCGCGGGGCGGGCGCGCCCGCGCGGGGCGGGGCGCGCGC 317  
QY 242 AGGTTGAACGAGAGGGCGTGGCCAGCAGCTGCGCAGGGGCGGAGAACCGCTGGGGGA 301  
Db 316 GGGGCGGGGCGCGCGGGGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 257  
QY 302 ACCCTTGCTCGCTCTCGCGCTCGCTCTAGGATCCCGCGAAAGAGACGCGGCGGAAAG 361  
Db 256 GGGGGGCGGGGCG 197  
QY 362 CGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGG 421  
Db 196 CGCGCGCGGGGCG 140  
QY 422 CTTGCTGTATGACTGCTGAGAGCTGCTTCCCGGCGGAGCCTTCTTCGAGGGTTC 481  
Db 139 CGGGGCG 83  
QY 482 ACGTCCAGGACCGGGGCTCGGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 82 CCCGCGGGNNGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24

RESULT 13  
US-10-425-115-2989  
; Sequence 2989, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 2989  
; LENGTH: 598  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(598)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_102723C.1  
US-10-425-115-2989

Query Match 5.3%; Score 51.6; DB 20; Length 598;  
Best Local Similarity 47.4%; Pred. No. 4.7e-05;  
Matches 108; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 73 GGCAGGAGGCCACAGCTGGAGCCTCAGCTCCGCGCGAGCCAGCGTGCCCTGTCTGTC 132  
Db 4 GCCCGGNGCGCGGGCG 63  
QY 133 CCGCGCTCCCGGGGCTTGGCTGCGCGCTCTGGAACCGCTGGGAGGACACGCGCGG 192  
Db 64 CNGCGGCGCGCGGGCG 123  
QY 193 AGGATGACAAAGTGTCTCGAGACATTTGCGGCGGGGCGGGGCGGGTGGCAGGGTGAAGC 252  
Db 124 CGGCGCGGCG 183  
QY 253 GAGGGGCGGTGGCGAGCTGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 184 GNNNGNNGCGCGGGGCG 231

RESULT 14  
US-10-437-963-49370/c  
; Sequence 49370, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Bouharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 49370  
; LENGTH: 763  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(763)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_51959C.1  
US-10-437-963-49370

Query Match 5.3%; Score 51.6; DB 19; Length 763;  
Best Local Similarity 46.4%; Pred. No. 5.1e-05;  
Matches 228; Conservative 0; Mismatches 259; Indels 4; Gaps 2;

QY 51 CTCCTCGCAGACTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 110  
Db 708 CTCGCGCGGATCTGGGGAGGGGAGGGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 649  
QY 111 GACCCAGCGTGCCTGTCTGTCTCCCGCGCTCCCGGGGCTTGGCTGCGCGCTCTTGACGCGG 170  
Db 648 GGCACACG 591  
QY 171 TGCGCAGCGGGACCAACCGCGGGAGGATGACGAAGGTGCTCGCGACATTTGCGCGCGCGG 230  
Db 590 GGGGGGCG 531  
QY 231 GGGCGCGTGGCAGGGTGGAAAGCGGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 290  
Db 530 GCGGCGNGCG 471  
QY 291 GGGCTGGGGAAACCCCTTGTGCTGTGCGCGCTCTGCGCGCTCTAGGATCC--CCGAAAAGGAGC 348  
Db 470 NCCGCGGGGCG 411  
QY 349 AGCGGCGGAAAGCG 408  
Db 410 GCGCGCGGGGCG 351  
QY 409 CTGCGCGCGCGCGCGCTGTCTATGACTCTGTGAGAGCTCTGCTTCCGCGCGGACCTT 468  
Db 350 CCGCGCGGGGCG 291  
QY 469 CTGCAAGGGTCCAAGTCCAGGACACCGCGCGGTTCGACACCCACACCCCGCGCGCGGACCC 528  
Db 290 GCG 231  
QY 529 TGGCTTGGGTG 539  
Db 230 GCGCGCGCGCG 220



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2005, 00:52:53 ; Search time 212 Seconds  
(without alignments)  
7563.921 Million cell updates/sec

Title: US-10-829-118-1

Perfect score: 980

Sequence: 1 ctcccgccggaacagcctg.....ctgcaccacggtaatgctga 980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*

2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*

3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*

4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*

5: /cgn2\_6/prodata/1/ina/6CTUS\_COMB.seq:\*

6: /cgn2\_6/prodata/1/ina/backfileseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	980	100.0	980	4	US-09-891-711-1	Sequence 1, Appli
2	980	100.0	2469	4	US-09-891-711-2	Sequence 2, Appli
C 3	966.4	98.6	67156	4	US-09-949-016-12284	Sequence 12284, A
C 4	966.4	98.6	67157	4	US-09-949-016-16258	Sequence 16258, A
C 5	126	12.9	4338	4	US-09-949-016-4816	Sequence 4816, Ap
C 6	124.4	12.7	4338	3	US-09-360-237-4	Sequence 4, Appli
C 7	124.4	12.7	4338	4	US-09-891-711-3	Sequence 3, Appli
C 8	124.4	12.7	4338	4	US-09-949-016-542	Sequence 542, App
C 9	53	5.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 10	52.2	5.3	300402	4	US-09-949-016-13632	Sequence 13632, A
C 11	48	4.9	42053	4	US-09-949-016-15924	Sequence 15924, A
C 12	47.4	4.8	154746	4	US-09-827-688-8	Sequence 8, Appli
C 13	47.4	4.8	154746	4	US-09-827-688-8	Sequence 8, Appli
C 14	46.8	4.8	2571	4	US-09-984-880-1	Sequence 1, Appli
C 15	46.8	4.8	2571	4	US-10-277-032-1	Sequence 14, Appl
C 16	46.8	4.8	20966	4	US-09-984-880-3	Sequence 3, Appli
C 17	46.8	4.8	20966	4	US-10-277-032-3	Sequence 3, Appli
C 18	46.6	4.8	1736	3	US-09-162-524-2	Sequence 2, Appli
C 19	46	4.7	867	4	US-09-902-540-9534	Sequence 9534, Ap
C 20	46	4.7	9867	4	US-09-902-540-1008	Sequence 1008, Ap
C 21	45.8	4.7	7218	1	US-08-232-463-14	Sequence 14, Appl
C 22	45.8	4.7	128516	4	US-09-949-016-13501	Sequence 13501, A
C 23	45.8	4.7	133157	4	US-09-949-016-12541	Sequence 12541, A
C 24	45.4	4.6	4335	3	US-08-974-549A-6	Sequence 6, Appli
C 25	45.4	4.6	4335	4	US-09-721-456-6	Sequence 6, Appli
C 26	45.2	4.6	4048	4	US-09-976-594-1053	Sequence 1053, Ap
C 27	45.2	4.6	122772	4	US-09-949-016-14132	Sequence 14132, A

28	44.8	4.6	28374	4	US-09-949-016-17508	Sequence 17508, A
C 29	44.4	4.5	4826	4	US-09-772-304A-1	Sequence 1, Appli
C 30	44.4	4.5	17098	4	US-09-949-016-15864	Sequence 15864, A
C 31	44.2	4.5	16321	4	US-09-949-016-12490	Sequence 12490, A
C 32	44.2	4.5	75212	4	US-09-949-016-13313	Sequence 13313, A
C 33	44.2	4.5	75212	4	US-09-949-016-13314	Sequence 13314, A
C 34	44.2	4.5	75212	4	US-09-949-016-13315	Sequence 13315, A
C 35	44.2	4.5	87523	4	US-09-949-016-12670	Sequence 12670, A
C 36	44.2	4.5	87523	4	US-09-949-016-15047	Sequence 15047, A
C 37	44.2	4.5	87523	4	US-09-949-016-15048	Sequence 15048, A
C 38	44.2	4.5	87523	4	US-09-949-016-15049	Sequence 15049, A
C 39	44.2	4.5	87869	4	US-09-949-016-11744	Sequence 11744, A
C 40	44.2	4.5	87869	4	US-09-949-016-15044	Sequence 15044, A
C 41	44.2	4.5	87869	4	US-09-949-016-15045	Sequence 15045, A
C 42	44.2	4.5	87869	4	US-09-949-016-15046	Sequence 15046, A
C 43	44	4.5	18197	4	US-09-949-016-12597	Sequence 12597, A
C 44	44	4.5	18197	4	US-09-949-016-15205	Sequence 15205, A
C 45	43.8	4.5	5437	1	US-07-661-610C-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-891-711-1  
; Sequence 1, Application US/09891711  
; Patent No. 6723553  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Promoter Sequences  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/09/891,711  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 980  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-891-711-1

Query Match 100.0%; Score 980; DB 4; Length 980;  
Best Local Similarity 100.0%; Pred. No. 3.1e-250;  
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CTCCGCGGGAACACGCTGGGCACCTCCATTCGGGCTGTTTACTCCCACTCTCGCGAG	60
Db	1	CTCCGCGGGAACACGCTGGGCACCTCCATTCGGGCTGTTTACTCCCACTCTCGCGAG	60
Qy	61	ACTGGCGGCGCGGCGGCGAGGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT	120
Db	61	ACTGGCGGCGCGGCGGCGAGGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT	120
Qy	121	GCCCTGTCTGTCCCGCGCTCCCGGGCTTCGCTGCGGCTTGTGAGACCCGTGGGACGG	180
Db	121	GCCCTGTCTGTCCCGCGCTCCCGGGCTTCGCTGCGGCTTGTGAGACCCGTGGGACGG	180
Qy	181	GACACCGCGGAGGATGACGAGGTGCTCCGACATTTGCGGCGGCGGGGCGCGGTGG	240
Db	181	GACACCGCGGAGGATGACGAGGTGCTCCGACATTTGCGGCGGCGGGGCGCGGTGG	240
Qy	241	CAGGGTGAAGCGGCGGCGCTGGCCAGCAGCTGCGAGCGCGGCGAGAACCGCTGGGGG	300
Db	241	CAGGGTGAAGCGGCGGCGCTGGCCAGCAGCTGCGAGCGCGGCGAGAACCGCTGGGGG	300
Qy	301	AACCCCTTGGTCCGCTCTGGCGCTCGCTCTAGGATCCCCGAAAAAGGAGCAGCGGCGG	360
Db	301	AACCCCTTGGTCCGCTCTGGCGCTCGCTCTAGGATCCCCGAAAAAGGAGCAGCGGCGG	360
Qy	361	GCGGCGAGCTGGCGGCGAGGATCTAGAAAGACTGCTGGGCGAGGCTCCCTGCCCCCGCG	420
Db	361	GCGGCGAGCTGGCGGCGAGGATCTAGAAAGACTGCTGGGCGAGGCTCCCTGCCCCCGCG	420

```

QY 421 GCCTGCTGTATGAGTCTGTGAGAGCTGCTTCCGGCGCGACCCCTTCCTGACGGGTC 480
DB 421 GCCTGCTGTATGAGTCTGTGAGAGCTGCTTCCGGCGCGACCCCTTCCTGACGGGTC 480
QY 481 CACGTCCAGGACCCGGGCTCGGACACCCACCCCGGCGGACCTGCGCTCGGTCG 540
DB 481 CACGTCCAGGACCCGGGCTCGGACACCCACCCCGGCGGACCTGCGCTCGGTCG 540
QY 541 CCCTTAAACCCGGGCGGTAGCTCGTTAAGATGCGGAAGTGTCCGGTCCCGAAACACCGAAA 600
DB 541 CCCTTAAACCCGGGCGGTAGCTCGTTAAGATGCGGAAGTGTCCGGTCCCGAAACACCGAAA 600
QY 601 CCCCCAAATCCCGCTTCCCGACCTCTGTACCCCGGCGCCACCGGACGACAGACTGGGCC 660
DB 601 CCCCCAAATCCCGCTTCCCGACCTCTGTACCCCGGCGCCACCGGACGACAGACTGGGCC 660
QY 661 TCCCGACCGGACCGGCTGCGGACACCCGCTGCGTGGGAACCGGAGACCTTTGTAAC 720
DB 661 TCCCGACCGGACCGGCTGCGGACACCCGCTGCGTGGGAACCGGAGACCTTTGTAAC 720
QY 721 GCACGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 780
DB 721 GCACGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 780
QY 781 TCCCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840
DB 781 TCCCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840
QY 841 TGGACACCGGACCTCCGTACGGCACTAGCAGGGGACTGAAAGGCTCTTCAGGTACTGC 900
DB 841 TGGACACCGGACCTCCGTACGGCACTAGCAGGGGACTGAAAGGCTCTTCAGGTACTGC 900
QY 901 TGGTGGGCGGTGATGGCTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTGACA 960
DB 901 TGGTGGGCGGTGATGGCTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTGACA 960
QY 961 CTGACACCGGTAATGCTGA 980
DB 961 CTGACACCGGTAATGCTGA 980

```

RESULT 2

```

US-09-891-711-2
; Sequence 2, Application US/09891711
; Patent No. 6723553
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-711-2

```

```

Query Match 100.0%; Score 980; DB 4; Length 2469;
Best Local Similarity 100.0%; Pred. No. 4.4e-250;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCGCGCGGGAACACGCTCGGCACTCCATTCGGGCTGTTTACTCCCAACTCTCGCGAG 60
DB 353 CTCGCGCGGGAACACGCTCGGCACTCCATTCGGGCTGTTTACTCCCAACTCTCGCGAG 412
QY 61 ACTGGCGCGCGGCGAGCGGCGCCACAGCTGAGGAGCTCAGCTCCGCGGACCCAGCGT 120
DB 413 ACTGGCGCGCGGCGAGCGGCGCCACAGCTGAGGAGCTCAGCTCCGCGGACCCAGCGT 472
QY 121 GCCTGTCTGCTCCGCGCTCCGCGGCTTGGTGGCGCTCTGGAGCGGCTGGGACGG 180

```

```

DB 473 GCCCTGTCTGCTCCCGCGCTCCCGGGCTTGGTGGCGGCTCTGGACGCGCTGGGCGAGCGG 532
QY 181 GACCAACCGCGGAGGATGGAAGAGTGTCTGCGGACATTTGCGGCGGCGGGGCGGCTGG 240
DB 533 GACCAACCGCGGAGGATGGAAGAGTGTCTGCGGACATTTGCGGCGGCGGGGCGGCTGG 592
QY 241 CAGGCTGAAGCGGAGGGGCGTGGCCAGGAGCTGCGGAGGCGGAGAACCGGCTGGGGG 300
DB 593 CAGGCTGAAGCGGAGGGGCGTGGCCAGGAGCTGCGGAGGCGGAGAACCGGCTGGGGG 652
QY 301 AACCTTCCGTCCGCTCTGCGGCTCTAGGATCCCGAAAAGAGACACGCGCGCGAAA 360
DB 653 AACCTTCCGTCCGCTCTGCGGCTCTAGGATCCCGAAAAGAGACACGCGCGCGAAA 712
QY 361 GCGGCAAGCTGGGCGGAGGATCTAGAAAGACTGCTTGGCGGCAAGGCTCCCTGCCCGCGGG 420
DB 713 GCGGCAAGCTGGGCGGAGGATCTAGAAAGACTGCTTGGCGGCAAGGCTCCCTGCCCGCGGG 772
QY 421 GCCTGCTGTGATGAGACTGTGAGAGCTGCTTCCGCGCGGACCTTCTCTGACAGGGTC 480
DB 773 GCCTGCTGTGATGAGACTGTGAGAGCTGCTTCCGCGCGGACCTTCTCTGACAGGGTC 832
QY 481 CACGTCCAGGACCCGGCGCTCGGACACCCACCCCGGCGGCGACCTGCGCTGGGTGC 540
DB 833 CACGTCCAGGACCCGGCGCTCGGACACCCACCCCGGCGGCGACCTGCGCTGGGTGC 892
QY 541 CCCTTAAACCCGGGCGGTAGCTGTTAAGATGCGGAAGTGTCCGGTCCCGAAACACCGGAAA 600
DB 893 CCCTTAAACCCGGGCGGTAGCTGTTAAGATGCGGAAGTGTCCGGTCCCGAAACACCGGAAA 952
QY 601 CCGGAAATCCCGCTTCCCGGACCTCTGACCCCGGCGGCGGACGACAGACTGGGCG 660
DB 953 CCGGAAATCCCGCTTCCCGGACCTCTGACCCCGGCGGCGGCGGACGACAGACTGGGCG 1012
QY 661 TCCCGACCGGACCGGCTGCGGACACCCGCTGCGTGGCGGAAACCGAGAGACCTTTGTAAC 720
DB 1013 TCCCGACCGGACCGGCTGCGGACACCCGCTGCGTGGCGGAAACCGAGGACCTTTGTAAC 1072
QY 721 GCCACGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 780
DB 1073 GCCACGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 1132
QY 781 TGCCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTTGTTCATGTAAA 840
DB 1133 TGCCGCTTAAAGATGTTGGGTGACTTAGATGCTAGGATCAGTTTGTTCATGTAAA 1192
QY 841 TGGACACCGGAGCTCCGTACGGCACTAGCAGGGGACTGAAAGGCTCTTCAGGTACTGC 900
DB 1193 TGGACACCGGAGCTCCGTACGGCACTAGCAGGGGACTGAAAGGCTCTTCAGGTACTGC 1252
QY 901 TGGTGGGCGGTGATGCGCTTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTGACA 960
DB 1253 TGGTGGGCGGTGATGCGCTTACAGGCGGATCAGACAGATTTGTGTCTTCTGGAACCTGACA 1312
QY 961 CTGACACCGGTAATGCTGA 980
DB 1313 CTGACACCGGTAATGCTGA 1332

```

RESULT 3

```

US-09-949-016-12284/c
; Sequence 12284, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

```



; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12284  
 ; LENGTH: 67156  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-12284

Query Match 98.6%; Score 966.4; DB 4; Length 67156;  
 Best Local Similarity 99.8%; Pred. No. 6.2e-246;  
 Matches 978; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 CTCGGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
 Db 2126 CTCGGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 2067  
 QY 61 ACTGGCGGCGGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCCT 120  
 Db 2066 ACTGGCGGCGGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCCT 2007  
 QY 121 GCCCTGTCTGTCCCGCGCTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGGAGCGG 180  
 Db 2006 GCCCTGTCTGTCCCGCGCTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGGAGCGG 1947  
 QY 181 GACCAAGCGGCGGAGGATGACCAAGGTGCTCCGACATTTGCGGCGGCGGCGGCGG 240  
 Db 1946 GACCAAGCGGCGGAGGATGACCAAGGT-CTCGGACATTTGCGGCGGCGGCGGCGG 1888  
 QY 241 CAGGGTGGAAAGCGGAGGCGCTGGCCAGCAGCTGCGAGCGCGGAGAACCGCTGGGG 300  
 Db 1887 CAGGGTGGAAAGCGGAGGCGCTGGCCAGCAGCTGCGAGCGCGGAGAACCGCTGGGG 1828  
 QY 301 AACCTTGTCTGCTTCCGCGCTCTGAGGATCCCGAAAGGAGCAGCGGCGCGAAA 360  
 Db 1827 AACCTTGTCTGCTTCCGCGCTCTGAGGATCCCGAAAGGAGCAGCGGCGCGAAA 1768  
 QY 361 GCGGCGGCGGCGGCGGAGCTAGAAAGACTGCTGCGGCGGAGGCTCCCTGCGCGCGCG 420  
 Db 1767 GCGGCGGCGGCGGCGGAGCTAGAAAGACTGCTGCGGCGGAGGCTCCCTGCGCGCGCG 1708  
 QY 421 GCTGTGTCTGCTGCTGAGAGCTGCTTCCCGCGGAGCCTTCTTCCGCGGCGGCTC 480  
 Db 1707 GCTGTGTCTGCTGCTGAGAGCTGCTTCCCGCGGAGCCTTCTTCCGCGGCGGCTC 1648  
 QY 481 CAGCTCCAGGCGGCGGCTCGGACACCCCGCGGCGGCGGCGGCGGCGGCGGCTG 540  
 Db 1647 CAGCTCCAGGCGGCGGCTCGGACACCCCGCGGCGGCGGCGGCGGCGGCGGCTG 1588  
 QY 541 CCCTTAAACCGGCGGCTAGCTGTTAAGATGGGGAAGTGTCCGCTCCGGAACACGCGAAA 600  
 Db 1587 CCCTTAAACCGGCGGCTAGCTGTTAAGATGGGGAAGTGTCCGCTCCGGAACACGCGAAA 1528  
 QY 601 CCCGAAATCCCGCTGCGGAGCTCTGACCCCGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 Db 1527 CCCGAAATCCCGCTGCGGAGCTCTGACCCCGCGGCGGCGGCGGCGGCGGCGGCGG 1468  
 QY 661 TCCCGAAGCGGCGGCTGCGGGAACACCGGTGCGTGGGAAAGGAGGAGGAGGAGGAGG 720  
 Db 1467 TCCCGAAGCGGCGGCTGCGGGAACACCGGTGCGTGGGAAAGGAGGAGGAGGAGGAG 1408  
 QY 721 GCGAGCTGTTGCTCTTTTGAAGAAACAAAGATAAATGTGTTAACTCTCTGAAAGCT 780  
 Db 1407 GCGAGCTGTTGCTCTTTTGAAGAAACAAAGATAAATGTGTTAACTCTCTGAAAGCT 1348  
 QY 781 TCGCGCTTAAAGATGCTGCGGAGCTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 840  
 Db 1347 TCGCGCTTAAAGATGCTGCGGAGCTTAGATGCTAGGATCAGTTGTTTCAATGTAAA 1288  
 QY 841 TGGACGAGCGGCGGCTCCGCTAGCGGAGCTAGCGGAGCTGAAAGCGCTCTCAGGTA 900

Db 1287 TGGACAGCGGCGGACTCCGCTACGGCACTAGAGGGGACTGAAAGCGCTCTTCAAGTACTGC 1228  
 QY 901 TGGTGGGGGCTGATCGCTACAGGCGGATCAGACAGTCTTGTCTCTTGGAACTTGACA 960  
 Db 1227 TGGTGGGGGCTGATCGCTACAGGCGGATCAGACAGTCTTGTCTCTTGGAACTTGACA 1168  
 QY 961 CTGCACCAAGCGTAAATGCTGA 980  
 Db 1167 CTGCACCAAGCGTAAATGCTGA 1148  
 RESULT 4  
 US-09-949-016-16558/c  
 ; Sequence 16558, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16558  
 ; LENGTH: 67157  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-16558

Query Match 98.6%; Score 966.4; DB 4; Length 67157;  
 Best Local Similarity 99.8%; Pred. No. 6.2e-246;  
 Matches 978; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 CTCGGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
 Db 2126 CTCGGCGGGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAG 2067  
 QY 61 ACTGGCGGCGGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCCT 120  
 Db 2066 ACTGGCGGCGGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCCT 2007  
 QY 121 GCCCTGTCTGTCCCGCGCTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGGAGCGG 180  
 Db 2006 GCCCTGTCTGTCCCGCGCTCCCGGGCTTTCGCTGCGCGCTCTGGAGCGCGGAGCGG 1947  
 QY 181 GACCAAGCGGCGGAGGATGACCAAGGTGCTCCGACATTTGCGGCGGCGGCGGCGGCTG 240  
 Db 1946 GACCAAGCGGCGGAGGATGACCAAGGT-CTCGGACATTTGCGGCGGCGGCGGCGGCTG 1888  
 QY 241 CAGGGTGGAAAGCGGAGGCGCTGGCCAGCAGCTGCGAGCGCGGAGAACCGCTGGGG 300  
 Db 1887 CAGGGTGGAAAGCGGAGGCGCTGGCCAGCAGCTGCGAGCGCGGAGAACCGCTGGGG 1828  
 QY 301 AACCTTGTCTGCTTCCGCGCTCTGAGGATCCCGAAAGGAGGAGGAGGAGGAGGAGG 360  
 Db 1827 AACCTTGTCTGCTTCCGCGCTCTGAGGATCCCGAAAGGAGGAGGAGGAGGAGGAGG 1768  
 QY 361 GCGGCGGCGGCGGCGGAGCTAGAAAGACTGCTTGGCGGCGGAGGCTCCCTGCGCGCGCG 420  
 Db 1767 GCGGCGGCGGCGGCGGAGCTAGAAAGACTGCTTGGCGGCGGAGGCTCCCTGCGCGCGCG 1708  
 QY 421 GCTGTGTCTGCTGCTGAGAGCTGCTTCCCGCGGAGCCTTCTTCCGCGGCGGCTC 480  
 Db 1707 GCTGTGTCTGCTGCTGAGAGCTGCTTCCCGCGGAGCCTTCTTCCGCGGCGGCTC 1648  
 QY 481 CAGCTCCAGGCGGCGGCTCGGACACCCCGCGGCGGCGGCGGCGGCGGCGGCTG 540

```
Db 1647 CACGTCACGCGCGCTCGGACACCCACACCCCGCGGCACTGCCCTGGGTGC 1588
QY 541 CCCTTAACCCGGCGGTAGCTGTTAAGATGCGAAGTGTCCGGTCCGGAACACCGGAAA 600
Db 1587 CCCTTAACCCGGCGGTAGCTGTTAAGATGCGAAGTGTCCGGTCCGGAACACCGGAAA 1528
QY 601 CCCCAATCCCGCTGCCGACCTCTGACCCCGGCCCCACGCGGACGACAGCTGGGCC 660
Db 1527 CCCCAATCCCGCTGCCGACCTCTGACCCCGGCCCCACGCGGACGACAGCTGGGCC 1468
QY 661 TCCCGACGCGACGCGCTGCCGCGACACCGGTGCGTGCAGAAACGAGAGACCTTTGTAAAC 720
Db 1467 TCCCGACGCGACGCGCTGCCGCGACACCGGTGCGTGCAGAAACGAGAGACCTTTGTAAAC 1408
QY 721 GCCAGTGTGTCTCTTTTGAAGAAACAAAGATTAATGTGTTAAACTGTCTGAAAGCT 780
Db 1407 GCCAGTGTGTCTCTTTTGAAGAAACAAAGATTAATGTGTTAAACTGTCTGAAAGCT 1348
QY 781 TGCGGCTAAAGATGTCGGGTGACTTAGATGCTAGGATCAGTTTGTTCATGTAA 840
Db 1347 TGCGGCTAAAGATGTCGGGTGACTTAGATGCTAGGATCAGTTTGTTCATGTAA 1288
QY 841 TGGACACGCGGACTCCGTACGCGCACTAGCAGGCGACTGAAAGCGTCTTCAGGTACTGC 900
Db 1287 TGGACACGCGGACTCCGTACGCGCACTAGCAGGCGACTGAAAGCGTCTTCAGGTACTGC 1228
QY 901 TGGTGGCGGTGATCGGTACAGGCGGATCAGACAGTTTGTCTTCGGAATTTGACA 960
Db 1227 TGGTGGCGGTGATCGGTACAGGCGGATCAGACAGTTTGTCTTCGGAATTTGACA 1168
QY 961 CTGACACGCGTAATGCTGA 980
Db 1167 CTGACACGCGTAATGCTGA 1148
```

```
RESULT 5
US-09-949-016-4816/c
; Sequence 4816, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4816
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4816
```

```
Query Match 12.9%; Score 126; DB 4; Length 4338;
Best Local Similarity 100.0%; Pred. No. 2.4e-23;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CTCGCGGGGAACACGCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGGGGAACACGCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGGCGCGGCGACGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 120
Db 66 ACTGGGCGCGGCGACGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 7
```

```
QY 121 GCCCTG 126
Db 6 GCCCTG 1
RESULT 6
US-09-360-237-4/c
; Sequence 4, Application US/09360237
; Patent No. 6322962
; GENERAL INFORMATION:
; APPLICANT: BROWN, MICHAEL S.
; APPLICANT: CHENG, DONG
; APPLICANT: ESPENSHADE, PETER J.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: RAWSON, ROBERT B.
; APPLICANT: SAKAI, JURO
; TITLE OF INVENTION: STEROID-REGULATED SITE-1 PROTEASE AND ASSAYS OF
; TITLE OF INVENTION: MODULATORS THEREOF
; FILE REFERENCE: UTXD:567
; CURRENT APPLICATION NUMBER: US/09/360,237
; CURRENT FILING DATE: 1999-07-23
; EARLIER APPLICATION NUMBER: 60/096,571
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Human
US-09-360-237-4
```

```
Query Match 12.7%; Score 124.4; DB 3; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.3e-23;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CTCGCGGGGAACACGCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGGGGAACACGCTGGGCACTCCATTCCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGGCGCGGCGGCGACGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 120
Db 66 ACTGGGCGGCGGCGGCGACGAGGCCACACAGCTGGAGGCTCAGCTCCGCGGACCCAGCGT 7
```

```
QY 121 GCCCTG 126
Db 6 GCCCTG 1
```

```
RESULT 7
US-09-891-711-3/c
; Sequence 3, Application US/09891711
; Patent No. 6723553
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (497)..(3655)
US-09-891-711-3
```

```
Query Match 12.7%; Score 124.4; DB 4; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.3e-23;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 CTCGCGGCGAACAAGCCCTGGGCACTCCATTCCGGGCTGTTTACTTCCCAACTCTCGCGAG 60  
Db |||||  
126 CTCGCGGCGAACAAGCCCTGGGCACTCCATTCCGGGCTGTTTACTTCCCAACTCTCGCGAG 67  
QY 61 ACTGGGCGCGGGCAGCGAGGCCACACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120  
Db |||||  
66 ACTGGGCGAGCGGGCCAGCGAGGCCACACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 7  
QY 121 GCCCTG 126  
Db 6 GCCCTG 1  
RESULT 8  
US-09-949-016-542/c  
; Sequence 542, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 542  
; LENGTH: 4338  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-542

Query Match 12.7%; Score 124.4; DB 4; Length 4338;  
Best Local Similarity 99.2%; Pred. No. 6.3e-23; Mismatches 1; Indels 0; Gaps 0;  
Matches 125; Conservative 0;  
QY 1 CTCGCGGCGAACAAGCCCTGGGCACTCCATTCCGGGCTGTTTACTTCCCAACTCTCGCGAG 60  
Db |||||  
126 CTCGCGGCGAACAAGCCCTGGGCACTCCATTCCGGGCTGTTTACTTCCCAACTCTCGCGAG 67  
QY 61 ACTGGGCGCGGGCAGCGAGGCCACACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120  
Db |||||  
66 ACTGGGCGAGCGGGCCAGCGAGGCCACACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 7  
QY 121 GCCCTG 126  
Db 6 GCCCTG 1

RESULT 9  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765

; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2  
Query Match 5.4%; Score 53; DB 3; Length 4403765;  
Best Local Similarity 48.4%; Pred. No. 0.0074;  
Matches 140; Conservative 2; Mismatches 147; Indels 0; Gaps 0;  
QY 48 CAATCTCTCGGAGACTGGGCGCGCGGCGGAGGATGACGAAGGTGCTCGCGACATTTGCCGCGG 107  
Db 1634333 CACCCGCGACCGGCGGCGGCGGCGGATCTGTGGGCAACCGTGGGCGG 1634274  
QY 108 GCGGACCCAGCGTGCCTCTGTCTCCCGGCTTCCGGGGCTTCCGTCGCGGCTCTCGACG 167  
Db 1634273 GCGGCTCGGCGCGCGCCGCGCAAGTCGCGGGCGCGCGGGGTGTTCGCGCAC 1634214  
QY 168 CGTGGGCGAGCGGCGGACCGCGCGGAGGATGACGAAGGTGCTCGCGACATTTGCCGCGG 227  
Db 1634213 CGGCGGCGCGGTGGGCGCGGTGGGCGCGCGCGCGCGCGCGGGGTAGCGCGG 1634154  
QY 228 CGGCGGCGCGGTGGGCGGAGGTTGGAAGCGGAGGCGGTGGCGAGCGAGTGCAGCGCGCGG 287  
Db 1634153 CTGGCTGTGGCAATGTTGGAGTCGGMGGGCGCGGCGGCGAGAGCTCTGCTGGYGGG 1634094  
QY 288 AACGCGCTGGGGGAACCCCTTGTCTCGCTCTGCGGCTCGCTCTAGGATCC 336  
Db 1634093 AACCGCGGCGCGCGCGCGCAACCGCGGACTGTTTCGGGGTTCGCGGGAACC 1634045

RESULT 10  
US-09-949-016-13632  
; Sequence 13632, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13632  
; LENGTH: 300402  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(300402)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13632

Query Match 5.3%; Score 52.2; DB 4; Length 300402;  
Best Local Similarity 48.8%; Pred. No. 0.0044;  
Matches 141; Conservative 0; Mismatches 148; Indels 0; Gaps 0;  
QY 151 CQTGGCGCTCTGGACCGCGTGGGCGGAGCCACCGGAGGATGACCAAGGTGCT 210  
Db |||||  
2216 CGGCGCGGAGGTGCGGAGCGGAGCTCGCGCTCGCGCGGACTGGGGACCCGCGAGCC 2275  
QY 211 CGCGACATTTGCGCGCGGCGGCGCGGTGGCGAGGTGGAAGCGGAGGGCGTGGCGAGCG 270  
Db |||||  
2276 TCGGGTGGCGCTGCAGGAGTCCGGGAGGCGCGGGGAGCGCGGGCGCTGCACCGCG 2335

QY 271 AGCTGCCAGCGCGAGAACCGCTGGGGAAACCCCTGGTCCGTCTGCGCGTCCGCTCTA 330  
Db 2336 CGCGCTCGGAGGTGGAGCGCGCGGTAGCCCGAGGGACAGGACAGAGGGATCGCGC 2395  
QY 331 GGATCCCGAAAGAGACAGCGGCGGAAAGGGCCAGGCTGGGCCAGGATCTAGAAGA 390  
Db 2396 GGATCCCGGCGGAGGGGTATCGGGTCCGGAACGCGCCGGGGCACCGGGCGGAGCGG 2455  
QY 391 CTGCTGCGCAGGCTCCCTGCGCCCGCGGCGCTGCTCATGGACTCG 439  
Db 2456 GGGCCCGCGCGGCTGATGCTGCCGAGCCCGGCTCGCTTGGAGGCG 2504

RESULT 11  
US-09-949-016-15924/c  
; Sequence 15924, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15924  
; LENGTH: 42053  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(42053)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15924

Query Match 4.9%; Score 48; DB 4; Length 42053;  
Best Local Similarity 44.9%; Pred. No. 0.028;  
Matches 183; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 21 GGCATCCATTCCGGGCTGTTTACTCCGAACTCTCCGAGACTGGCGGCGCGGCGAGG 80  
Db 1181 GGCCTTTGCGGAGCGCGGATTCCTCCAACTTCTCCCTCGCTCCCGCGCGGGGA 1122  
QY 81 AGGCCACAGCTGGGAGCTCAGCTCCGCGGACCCAGCGTGCCTGTCTGTCCCGGCTC 140  
Db 1121 AGAGGGACAGCGCGGCGCCCGCAGACCCGAGGGCTCGGCGCCCGCCGACCCGACCC 1062  
QY 141 CGCGGGCTTTCGTGCGCGCTCTGAGCGCGTGGGAGCGGGAACACCGCGGAGGATGA 200  
Db 1061 CTCGGCTTTTCGGCGGCGCGAGCGCCCTCAGCGGGAGCGCCGAGGAGGAAGAA 1002  
QY 201 CGAAGGTGCTCGGACATTTTCGGCGGCGGCGCGGCTGGCAGGTGGAGCGGAGGGC 260  
Db 1001 AGAGGGCGAGGAGGAGAGGGGGAGCTTCGAGACAGAAAGCAGGAGGAGGCGGCG 942  
QY 261 GTGGCCAGCGAGTGCAGCGCGCGAGAACCGGCTGGGGGAAACCTTGTCTCGCTCTGG 320  
Db 941 GCGGGCGGGAACCGCGGGGGCGCTTCAGCAACGCGGCGAGTGACAGCGCCCGCCGGA 882  
QY 321 CGTGCCTTAGGATCCCCGAAAGGAGCAAGCGCGCGAAAGCGGCGAGGCTGGCGAGA 380  
Db 881 CCGCGCGCGCGGAGCCCGGAGAGAAACGGGGGCTCGGGGAGAAAGCGGGTCCGGG 822  
QY 381 TCTAGAAAGACTGCTGGCGAGGCTCCCTGCCCCCGCGGCTGCTG 428

Db 821 GGAGAAACGGGGCGGAGGAGGCGCGGAGCGCGGGGAGAGCCG 774

RESULT 12  
US-09-827-688-8  
; Sequence 8, Application US/09827688  
; Patent No. 6821955  
; GENERAL INFORMATION:  
; APPLICANT: ORSON, FRANK  
; APPLICANT: KINSEY, BERMA  
; APPLICANT: BHOGAL, BALBIR  
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION  
; TITLE OF INVENTION: AGENTS  
; FILE REFERENCE: P01949US1/10004014  
; CURRENT APPLICATION NUMBER: US/09/827,688  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,680  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 154746  
; TYPE: DNA  
; ORGANISM: HERPESVIRUS 2  
US-09-827-688-8

Query Match 4.8%; Score 47.4; DB 4; Length 154746;  
Best Local Similarity 49.8%; Pred. No. 0.065;  
Matches 120; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 60 GACTGGGGCGCGGGCCAGCGAGGCCCCACAGCTGGAGCCTCAGCTCCGCGACCCAGCG 119  
Db 133282 GCGCGGGGACGCGGCGCGGGGAGCGGCGCGGGGGGACGCGGCGCGGGGGACG 133341  
QY 120 TGCCCTGTCTCTCCCGCGCTCCCGGGGCTTTCGTCGCGCTCTGGACCGCCCTGGGACG 179  
Db 133342 GACGCGGCGCGGGGAGCGGCGCGGGGGAGCGGCGCGGGGGCGCGGGGGCGG 133401  
QY 180 GGACACCGCGGAGGATGAGCAAGTGTCTCGGACATTTTCGCGCGCGGGGGCGGTG 239  
Db 133402 GGGGCGCGGGGAGCGGGGGAGCGGGGGAGCGGGGGGACGCGGGGGGACGCGGG 133461  
QY 240 GCAGGTGGAAGCGGAGGGGCTGCGCAGCGAGCTGCCAGCGCGGAGAACCGCTGGG 299  
Db 133462 GACGCGGGGAGCGGGGGAGCGGGGGAGCGGGGGGACGCGGGGGGACGCGGGG 133521  
QY 300 G 300  
Db 133522 G 133522

RESULT 13  
US-09-827-688-8/c  
; Sequence 8, Application US/09827688  
; Patent No. 6821955  
; GENERAL INFORMATION:  
; APPLICANT: ORSON, FRANK  
; APPLICANT: KINSEY, BERMA  
; APPLICANT: BHOGAL, BALBIR  
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION  
; TITLE OF INVENTION: AGENTS  
; FILE REFERENCE: P01949US1/10004014  
; CURRENT APPLICATION NUMBER: US/09/827,688  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,680  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 154746  
; TYPE: DNA  
; ORGANISM: HERPESVIRUS 2  
US-09-827-688-8



***This Page Blank (uspiv,***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2005, 00:48:23 ; Search time 3169 Seconds  
(without alignments)  
11771.216 Million cell updates/sec

Title: US-10-829-118-1

Perfect score: 980

Sequence: 1 ctccgcggcgaacacgcctg.....ctgcaccacgtaagtctga 980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_gss1:\*

9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	471.6	48.1	500	6	CB049709 NISC gj12
C 2	468	47.8	468	1	AI792242 ov05e09.y
C 3	451.8	46.1	455	1	AI760651 w166g09.x
C 4	448	45.7	450	2	AW051661 wx27b06.x
C 5	440.6	45.0	1240	5	BQ437878 AGENCOURT
C 6	435	44.4	435	1	AI685715 tu37c07.x
C 7	435	44.4	435	1	AI733521 ov05e09.x
C 8	435	44.4	435	2	BF446188 7p33f05.x
C 9	432	44.1	432	1	AI554846 te45b04.x
C 10	430.2	43.9	436	2	AW005919 wz90e09.x
C 11	411.8	42.0	415	1	AI362974 qv87h07.x
C 12	404	41.2	404	2	BE048839 hr54e04.x
C 13	362.8	37.0	399	1	AI498080 tm66a04.x
C 14	303.6	31.0	457	2	BF570483 60218e293
C 15	263	26.8	283	1	AI198389 q162f10.x
C 16	249	25.4	251	5	BX100137 BX100137
C 17	163	16.6	582	5	BP364386 BP364386
C 18	159.6	16.3	567	7	CR537300 DKFZp459P
C 19	141.8	14.5	584	5	BP358100 BP358100
C 20	137.6	14.0	742	7	CR790685 DKFZp4590
C 21	133.6	13.6	649	7	CR750317 DKFZp469J
C 22	130.2	13.3	390	7	CR549288 DKFZp4590
C 23	130.2	13.3	536	7	CR549308 DKFZp459P
C 24	128.8	13.1	575	1	AL701753 DKFZp6860

C 25	127.2	13.0	582	5	BP357251
C 26	125	12.8	701	6	CD617566 56088007H
C 27	123.4	12.6	249	4	BM146635 TCAAP1E64
C 28	121.8	12.4	564	1	BX957183 DKFZp781L
C 29	120.2	12.3	144	1	AA998897 ov05e09.s
C 30	119.2	12.2	350	7	CR545379 DKFZp459M
C 31	116.4	11.9	990	5	EX443626 BX443626
C 32	112.4	11.5	506	4	BM194303 TCAAP1Q84
C 33	112.2	11.4	501	7	CR556966 DKFZp459A
C 34	110.2	11.2	829	4	BM553744 AGENCOURT
C 35	105.8	10.8	357	5	BU431331 UI-HF-BN0
C 36	105.8	10.8	505	5	BU431330 UI-HF-BN0
C 37	104.6	10.7	569	1	AL701734 DKFZp686K
C 38	92	9.4	596	4	BI545503 603187665
C 39	90.4	9.2	780	4	BG702439 602683790
C 40	90.2	9.2	454	1	AA045615 zk83a12.r
C 41	82.6	8.4	932	9	AL066742 Drosophil
C 42	75.2	7.7	1300	9	AG430756 Mus muscu
C 43	74	7.6	581	5	BP222057 BP222057
C 44	73.4	7.5	879	9	AG130935 Pan trogl
C 45	72.2	7.4	1165	9	AG030649 Pan trogl

## ALIGNMENTS

RESULT 1  
CB049709/c

LOCUS

DEFINITION

3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CB049709 500 bp mRNA linear EST 17-JAN-2003  
NISC gj12f01.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:3271441

3', mRNA sequence.

CB049709

CB049709.1 GI:27787996

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 500)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

CDNA Library Preparation:

DNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[info@image.llnl.gov](mailto:info@image.llnl.gov)

Plate: LLAM8008 row: L column: 2

Seq primer: -21M13 forward primer (ABI).

Location/Qualifiers

1. 500

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3271441"

/sex="male"

/dev\_stage="adult"

/clone\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Pr28"

/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)

with a modified polylinker; Plasmid DNA from the

normalized library NCI CGAP Pr22 was prepared, and ss

circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from a pool

of 5,000 clones made from the same library (clonoids

985608-986759, 110192-110195, and 1217928-1220615).

Subtraction by Bento Soares and M. Fatima Bonaldo. "





Qy	624	TCTGTACCCCGCGCCCCACCGGACGACAGACTGGGCCCTCCCGACGCGCAGCGCTGCGG	683
Db	150	TCCTGACCCCGCGCCCCACCGGACGACAGACTGGGCCCTCCCGACGCGCAGCGCTGCGG	91
Qy	684	GGACACGGTGGCGTGGCAACCGGAGGACCTTGTAAAGCCACGCTGTGTGCTCTTTTGA	743
Db	90	GGACACGGTGGCGTGGCAACCGGAGGACCTTGTAAAGCCACGCTGTGTGCTCTTTTGA	31
Qy	744	AAACAGAATAAATGTGTAAACTGTC	771
Db	30	AAACAGAATAAATGTGTAAACTGTC	3
RESULT 5			
LOCUS	BQ437878	1240 bp	mRNA linear EST 24-MAY-2002
DEFINITION	AGENCOURT_7990066 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6176570		
ACCESSION	BQ437878		
VERSION	BQ437878.1	GI:21176954	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1240)		
TITLE	NTH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: ATCC		
FEATURES	source		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Plate: L1AM13553 Row: c Column: 03		
	High quality sequence stop: 421.		
	Location/Qualifiers		
	1..1240		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6176570"		
	/tissue_type="leiomyosarcoma"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH MGC 71"		
	/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."		
ORIGIN			
Query Match	45.0%	Score 440.6; DB 5; Length 1240;	
Best Local Similarity	94.4%	Pred. No. 1.4e-97;	
Matches 490; Conservative	0; Mismatches 24; Indels 5; Gaps 3;		
Qy	274	TGCCAGCGCGGAGAACCGCTGGGGGAAACCTTGGTCCGCTCTGCGCGTCTCTAGGA	333
Db	1	TGCCAGCGCGGAGAACCGCTGGGGGAAACCTTGGTCCGCTCTGCGCGTCTCTAGGA	60
Qy	334	TCCCGAAGAGGACGCGGCGCGAAGCGCGCAGGCTGGGCCAGGATCTAGAAGACTG	393
Db	61	TCCCGAAGAGGACGCGGCGCGAAGCGCGCAGGCTGGGCCAGGATCTAGAAGACTG	120
Qy	394	CTGGCGCAGGCTCCCTGCCCGCGGCGCTGTGTATGACTCTGTGAGAGCTCGCTT	453
Db	121	CTGGCGCAGGCTCCCTGCCCGCGGCGCTGTGTATGACTCTGTGAGAGCTCGCTT	180
Qy	454	CCCGCGGAGCCCTTCTTCGAGGGGTCCACGTCCAGGACACCGCGGCTCGGACACCCAC	513
Db	181	CCCGCGGAGCCCTTCTTCGAGGGGTCCACGTCCAGGACACCGCGGCTCGGACACCCAC	240



Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation by: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
cDNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 435  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3647745"  
/sex="male"  
/dev stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="NCI-CGAP Pr28"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI CGAP Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonoids  
985608-986759, 110192-110195, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

# FEATURES

source

## ORIGIN

Query Match 44.4%; Score 435; DB 2; Length 435;  
Best Local Similarity 100.0%; Pred. No. 3e-96;  
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 337 CGAAGAAGGACGACGGCGGCGAAGCGCGAGCTCGGCGAGGATCTAGAAAGACTGCCT 396  
Db 435 CGAAGAAGGACGACGGCGGCGAAGCGCGAGCTCGGCGAGGATCTAGAAAGACTGCCT 376  
QY 397 GCGCAGGCTCCCTGCGCCCGCGGGGCTGCTGTATGAGTCTGTGGAGAGCTCGCTTCCC 456  
Db 375 GCGCAGGCTCCCTGCGCCCGCGGGGCTGCTGTATGAGTCTGTGGAGAGCTCGCTTCCC 316  
QY 457 GCGCGAGCTTCTTGAGGGGTTCAGGTCAGGACACGGCGGCTCGGACACCCACCC 516  
Db 315 GCGCGAGCTTCTTGAGGGGTTCAGGTCAGGACACGGCGGCTCGGACACCCACCC 256  
QY 517 GCGCGGGACCTGCGCTGGTCCCTTAACCCGGCGGTAGCTCGTTAAGATGCGAA 576  
Db 255 GCGCGGGACCTGCGCTGGTCCCTTAACCCGGCGGTAGCTCGTTAAGATGCGAA 196  
QY 577 GTGTCCGGTCCGGAACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGG 636  
Db 195 GTGTCCGGTCCGGAACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGG 136  
QY 637 CCCCACGGGACGACAGCTGGGCTCCCGACGCGAGCGGCTGCGGCGGACACCGGTGG 696  
Db 135 CCCCACGGGACGACAGCTGGGCTCCCGACGCGAGCGGCTGCGGCGGACACCGGTGG 76  
QY 697 TCGGAACCGGAGGACCTTTGTAACGCGACGTGTTTCTCTTTTGAAGAAACGAATAA 756  
Db 75 TCGGAACCGGAGGACCTTTGTAACGCGACGTGTTTCTCTTTTGAAGAAACGAATAA 16  
QY 757 ATGTGTTAAACTGTC 771  
Db 15 ATGTGTTAAACTGTC 1

RESULT 9  
AI554846/c  
LOCUS  
DEFINITION  
3', mRNA sequence.  
AI554846  
ACCSSION  
VERSION  
AI554846.1 GI:4487209  
432 bp mRNA linear EST 23-MAR-1999  
ta45b04.x1 Soares\_NhMPu\_S1 Homo sapiens cDNA clone IMAGE:2089615

## KEYWORDS

SOURCE  
ORGANISM

EST. Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 432)

## REFERENCE

AUTHORS

TITLE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 430.

Location/Qualifiers  
1. 432

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2089615"  
/tissue type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/clone\_lib="Soares\_NhMPu\_S1"  
/note="Organ: mixed (see below); Vector: pT7T3D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not 1;  
Site 2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBHM, pregnant uterus  
NBHPU, and fetal heart NBH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

Query Match 44.1%; Score 432; DB 1; Length 432;  
Best Local Similarity 100.0%; Pred. No. 1.6e-95;  
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 AAAAGAGACACGGGCGGCGAAGCGCGAGCTGGGCGAGGATCTAGAAAGACTGCCTGCG 399  
Db 432 AAAAGAGACACGGGCGGCGAAGCGCGAGCTGGGCGAGGATCTAGAAAGACTGCCTGCG 373  
QY 400 GCAGGCTCCCTGCGCCCGCGGCGCTGCTCATGAGTCTGTGGAGAGCTCGCTTCCCGCG 459  
Db 372 GCAGGCTCCCTGCGCCCGCGGCGCTGCTCATGAGTCTGTGGAGAGCTCGCTTCCCGCG 313  
QY 460 CCGACCTTCTCTGAGGGGTCCACGTCAGGACACCGCGGCTCGGACACCCACCCCGG 519  
Db 312 CCGACCTTCTCTGAGGGGTCCACGTCAGGACACCGCGGCTCGGACACCCACCCCGG 253  
QY 520 CCGGCGACCTGCGCTGGGTGCCCTTTAACCCGGGCGGTAGCTCGTTAAGATGCGAAGTG 579  
Db 252 CCGGCGACCTGCGCTGGGTGCCCTTTAACCCGGGCGGTAGCTCGTTAAGATGCGAAGTG 193  
QY 580 TCGGGTCCGGACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGGCGCC 639  
Db 192 TCGGGTCCGGACACGCGAAACCCCAATCCCGCTGCGGACCTCTGACCCCGGCGCC 133  
QY 640 CACGGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGGGACACCGGTGCGTGC 699  
Db 132 CACGGGACGACAGACTGGGCTCCCGACGCGAGCGGCTGCGGGACACCGGTGCGTGC 73  
QY 700 GAAACGAGGACTTTGTAAACGCGCAGTGTGCTCTTTTGAAGAAACGAATAAATG 759  
Db 72 GAAACGAGGACTTTGTAAACGCGCAGTGTGCTCTTTTGAAGAAACGAATAAATG 13  
QY 760 TGTGTTAACTGTC 771  
Db 760 TGTGTTAACTGTC 771



```

Db      12  TGTAAACTGTC 1

RESULT 10
AW005919/c
LOCUS   w290609.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2566120 3',
DEFINITION mRNA sequence.
ACCESSION AW005919
VERSION   AW005919.1 GI:5854686
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE   1 (bases 1 to 436)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(JCGRP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 541 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
FEATURES             source
    1..436
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:2566120"
        /tissue_type="anaplastic oligodendroglioma"
        /lab_host="DH10B"
        /notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
        modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
        strand cDNA was primed with a Not I - oligo(dT) primer [5'
        TGTTACCAATCTGAAGTGGAGCGCGCCGATAGGTTTTTTTTTTTTTTTTTTT
        T 3']; double-stranded cDNA was ligated to Eco RI
        adaptors (Pharmacia), digested with Not I and cloned into
        the Not I and Eco RI sites of the modified pT7T3 vector.
        Library is normalized, and was constructed by Bento
        Soares and M.Fatima Bonaldo."

ORIGIN
Query Match      43.9%; Score 430.2; DB 2; Length 436;
Best Local Similarity 99.3%; Pred. No. 4.5e-95;
Matches 432; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 343 AGAGACACGGCGCGAAAGCGCGCAGCTGGCGCAGGATCTAGAAAGCTGCTGCGCGCA 402
Db   |
QY 436 AGAGACACGGCGCGAAAGCGCGCAGCTGGCGCAGGATCTAGAAAGCTGCTGCGCGCA 377
Db   |
QY 403 GGTCTCTCCCGCGCGCGCTGCTCATGACTCGTGGAGACTCGCTTCCCGCGCG 462
Db   |
QY 376 GGTCTCTCCCGCGCGCGCTGCTCATGACTCGTGGAGACTCGCTTCCCGCGCG 317
Db   |
QY 463 ACCCTTCTCTCGAGGGGTCCACAGTCCAGGACACCGCGCGCTCGGACACCCACCGCGCG 522
Db   |
QY 316 ACCCTTCTCTCGAGGGGTCCACAGTCCAGGACACCGCGCGCTCGGACACCCACCGCGCG 257
Db   |
QY 523 GGCACCTGCGCTGGGTGCGCCCTTAACCCGGGGGTAGCTCGTTAAGATGCGGAAGTGTC 582
Db   |
QY 256 GGCACCTGCGCTGGGTGCGCCCTTAACCCGGGGGTAGCTCGTTAAGATGCGGAAGTGTC 197

```

363 GGCAGGCTGGCCAGGATCTAGAAAGACTGCTGGCGCAGAGCTCCCTGGCCCCCGGGGC 422  
Db |||||  
415 GGCAGGCTGGCCAGGATCTAGAAAGACTGCTGGCGCAGAGCTCCCTGGCCCCCGGGGC 356  
Qy |||||  
423 CTGCTGTATGACACTCGTGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGACAGGGGTCCA 482  
Db |||||  
355 CTGCTGTATGACACTCGTGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGACAGGGGTCCA 296  
Qy |||||  
483 CGTCAGGACACCGGGGGCTCGGACACCCACCCCGCGCGGCACTGCCCCCTGGGTGCC 542  
Db |||||  
295 CGTCAGGACACCGGGGGCTCGGACACCCACCCCGCGCGGCACTGCCCCCTGGGTGCC 236  
Qy |||||  
543 CTTAAACCCGGGCGGTAGCTCGTTAAGATGGGAAAGTCCGGTCCGGAACACGCGAAACC 602  
Db |||||  
235 CTTAAACCCGGGCGGTAGCTCGTTAAGATGGGAAAGTCCGGTCCGGAACACGCGAAACC 176  
Qy |||||  
603 CCAATCCCGCTGCCGACCTCCTGACCCCGCGGCCACCGGACGACAGCTGGGCCCTC 662  
Db |||||  
175 CCAATCCCGCTGCCGACCTCCTGACCCCGCGGCCACCGGACGACAGCTGGGCCCTC 116  
Qy |||||  
663 CGAGCGCAGCGCTGCCGGACACCGGTGCGTGGGAAACCGAGGACCTTTGTAAACGC 722  
Db |||||  
115 CGAGCGCAGCGCTGCCGGACACCGGTGCGTGGGAAACCGAGGACCTTTGTAAACGC 56  
Qy |||||  
723 CACGTGTTGCTCTTTTGAAGAAACCAAGAAATAAATGTTTAAACTGTCAAAA 777  
Db |||||  
55 CACGTGTTGCTCTTTTGAAGAAACCAAGAAATAAATGTTTAAACTGTCAAAA 1

RESULT 12  
LOCUS BE048839/c  
DEFINITION hr54e04.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3132318 3',  
mRNA sequence.  
ACCESSION BE048839  
VERSION BE048839.1 GI:8365894  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 404)  
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
[info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. .404  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3132318"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site: 1: Not 1; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library

FEATURES  
source

## ORIGIN

Query Match 41.2%; Score 404; DB 2; Length 404;  
Best Local Similarity 100.0%; Pred. No. 1.3e-88;  
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 370 CTGGCCAGGATCTAGAAAGACTGCTGGCGCAGAGCTCCCTGGCCCCCGGGGCCTGCTGT 429  
Db |||||  
404 CTGGCCAGGATCTAGAAAGACTGCTGGCGCAGAGCTCCCTGGCCCCCGGGGCCTGCTGT 345  
Qy 430 CATGACACTCGTGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGACAGGGGTCCACGTCCAG 489  
Db |||||  
344 CATGACACTCGTGAGAGCTCGCTTCCCGCGGAGCCCTTCCTGACAGGGGTCCACGTCCAG 285  
Qy 490 GCACCGCGGGCTCGGACACCCACCCCGCGCGGCACCTGCCCTGGGTGCCCTTTAAACC 549  
Db |||||  
284 GCACCGCGGGCTCGGACACCCACCCCGCGCGGCACCTGCCCTGGGTGCCCTTTAAACC 225  
Qy 550 CGGGCGGTAGCTCGTTAAGATGGGAAAGTGTCCGGTCCGGAACACCGCGAAACCCCAATC 609  
Db |||||  
224 CGGGCGGTAGCTCGTTAAGATGGGAAAGTGTCCGGTCCGGAACACCGCGAAACCCCAATC 165  
Qy 610 CGGCTGCCGACCTCCTGACCCCGCGGCCACCGGACGACAGCTGGGCTCCGACGC 669  
Db |||||  
164 CGGCTGCCGACCTCCTGACCCCGCGGCCACCGGACGACAGCTGGGCTCCGACGC 105  
Qy 670 GCAGCGCGCTGCCGGACACCGGTGCGTGGGAAACCGAGGACCTTTGTAAACCGCAGCTGT 729  
Db |||||  
104 GCAGCGCGCTGCCGGACACCGGTGCGTGGGAAACCGAGGACCTTTGTAAACCGCAGCTGT 45  
Qy 730 TTGCTCTTTTGAAGAAACCAAGAAATAAATGTTTAAACTGTCTG 773  
Db |||||  
44 TTGCTCTTTTGAAGAAACCAAGAAATAAATGTTTAAACTGTCTG 1

RESULT 13  
LOCUS AI498080/c

DEFINITION tm66a04.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2163054 3',  
mRNA sequence.  
ACCESSION AI498080  
VERSION AI498080.1 GI:4390062  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 389)  
REFERENCE NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute / National Institute of Neurological  
TITLE Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/STGAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 516 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 378.  
Location/Qualifiers  
1. .389  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2163054"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Brn25"  
/note="Organ: Brain; Vector: pTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCAATCTGAAGTGGAGCGCGCATAGTTTATTTTATTTTATTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

## ORIGIN

Query Match 37.0%; Score 362.8; DB 1; Length 389;  
Best Local Similarity 99.2%; Pred. No. 1.8e-78;  
Matches 375; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 400 CGAGGCTCCCTGCGCGCGGCTGCTGTCATGGACTCGTGAGAGCTCGCTTCCCGCG 459  
DB 389 CGAGGCTCCCTGCGCGCGGCTGCTGTCATGGACTCGTGAGAGCTCGCTTCCCGCG 330  
QY 460 CGGACCCCTTCTGCGAGGGTCCACGTCACAGCACCGCGGCTCGGACACCCACCCCGCG 519  
DB 329 CGGACCCCTTCTGCGAGGGTCCACGTCACAGCACCGCGGCTCGGACA-CCACCCCGCG 271  
QY 520 CGGGGACCTGCGCTGGTGGTCCCTTAAACCGGGCGGTAGCTGTTAAGATGCGGAAGTG 579  
DB 270 CGGGGACCTGCGCTGGTGGTCCCTTAAACCGGGCGGTAGCTGTTAAGATGCGGAAGTG 211  
QY 580 TCGGTCCGACACGCGAAACCCAAATCCCGCTGCCGACCTCTGACCCCGCGCC 639  
DB 210 TCGGTCCGACACGCGAAACCCAAATCCCGCTGCCGACCTCTGACCCCGCGCC 151  
QY 640 CACGGACGACAGACTGGGCTCCCGACGCGAGCGCTCGCGGACACCGGTGGTGC 699  
DB 150 CACGGACGACAGACTGGGCTCCCGACGCGAGCGCTCGCGGACACCGGTGGTGC 91  
QY 700 GAAACGAGGACTTTGTAAACGCGCTGTTGCTCTTTTGAAGAAACAAAGTAAATG 759  
DB 90 GAAACGAGGACTTTGTAAACGCGCTGTTGCTCTTTTGAAGAAACAAAGTAAATG 31  
QY 760 TGTAAACTGTCGAAA 777  
DB 30 TGTAAACTGTCAAAAA 13

RESULT 14  
BF570483/c  
LOCUS  
DEFINITION  
602186293T1 NTH\_MGC\_45 Homo sapiens cDNA clone IMAGE:4310518 3',  
mRNA sequence.  
ACCESSION  
BF570483  
VERSION  
BF570483.1 GI:11644195  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 457)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Linehan  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1185 row: j column: 23  
High quality sequence start: 30  
High quality sequence stop: 448.  
Location/Qualifiers

FEATURES  
source

1..457  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4310518"  
/tissue\_type="renal carcinoma (ascites)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 45"  
/note="Organ: Kidney; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 31.0%; Score 303.6; DB 2; Length 457;  
Best Local Similarity 91.7%; Pred. No. 6.9e-64;  
Matches 378; Conservative 0; Mismatches 24; Indels 10; Gaps 5;  
QY 283 GCGAAGACGCGTGGGGGACCCCTTGGTCCGCTCTGCGGCTCGCTTAGAGTCCCGAAA 342  
DB 447 GAGAACGCGCATGGGGGAAACCATTTGGTCCGCTCTGCGGCTCGCTTAGAGTCCCGAAA 388  
QY 343 AGGAGCAGCGGCGCGAAGCGGCCAGGCTGGGCCAGGATCTAGAAAGACTGCCT-GGCGC 401  
DB 387 AGGAGCAGCGGCGCGAAGCGGCCAGGCTGGGCCAGGATCTAGAAAGACTGCCTAGCGCG 328  
QY 402 AGGCTCCCTGCGCCCGCGCGGCC----TGCTGTCTAGGACT--CGTGGAGAGCTCGCTTCC 455  
DB 327 AGGCTCCCTGCGCCCGCGCGGCCCTAGCTCGATCATGGACTACGTCGGAGAGCTCGCTTCC 268  
QY 456 CGCGCGGACCTTCC--TGCAGGGTCCAGCTCCAGGCAACCGCGGCTCGGACACCCAC 513  
DB 267 CGCGCGGACCTTCCATGTCAGGAGGTCCACGTCAGGCAACCGCGGCTCGGACACCCAC 208  
QY 514 CCCCGCGCGGACCTGCGCTCGGTCGCC----TTAACCCGGCGGTAGCTCGTTAAGATG 572  
DB 207 CCCCGCGCGGACCTGCGCTCGGTCGCCCATTAACCCGGCGGTAGCTCGTTAAGATG 148  
QY 573 CGAAGTGTCCGTCGCGAACACGCGAAGCCCAATCCCGCTGCCCGACCTCTGACCC 632  
DB 147 CGAAGTGTCCGTCGCGAACACGCGAAGCCCAATCCCGCTGCCCGACCTCTGACCC 88  
QY 633 CGCGGCCACCGGAGCAGACAGCTGGGCTTCCGAGCGCGGAGCGCTGCGCG 684  
DB 87 CGCGGCCACCGGAGCAGACAGCTGGGCTTCCCGAGCGCGGCGCTGCGCG 36

## RESULT 15

AL198389/c  
LOCUS  
DEFINITION  
qi62f10.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:1861099 3',  
mRNA sequence.  
ACCESSION  
AL198389  
VERSION  
AL198389.1 GI:3750995  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 283)  
NCI/NIHDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project

JOURNAL COMMENT  
(CGAP/BTGP), Tumor Gene Index  
Unpublished (1998)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 443 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 238.

FEATURES  
source

1..263  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1861099"  
/issue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Brn25"  
/note="Organ: brain; Vector: pTT3D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCATCTGAGTGGGAGCGCGCATAGGTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
Library is normalized, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 26.8%; Score 263; DB 1; Length 263;  
Best Local Similarity 100.0%; Pred. No. 6.3e-54;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 521 CGGGACCTGCCCTGGTGGTCCCTTAACCGGGCGGTAGCTCGTTAAGATGGGAAGTGT 580  
DB 263 CGGGACCTGCCCTGGTGGTCCCTTAACCGGGCGGTAGCTCGTTAAGATGGGAAGTGT 204  
QY 581 CGGTCCGGAACACACGGAACCCCAATCCCGCTGCCCGACCTCTGTACCCCGGCCCC 640  
DB 203 CGGTCCGGAACACGGAACCCCAATCCCGCTGCCCGACCTCTGTACCCCGGCCCC 144  
QY 641 ACGGACGACAGACTGGGCTCCCGACGCGCGCTGCGGGACACCGGTGCGTGG 700  
DB 143 ACGGACGACAGACTGGGCTCCCGACGCGCGCTGCGGGACACCGGTGCGTGG 84  
QY 701 AACGAGGACCTTTTAAAGCCGCTGTCTCTTTTGAATAAAGATTAATGT 760  
DB 83 AACGAGGACCTTTTAAAGCCGCTGTCTCTTTTGAATAAAGATTAATGT 24  
QY 761 GTTAAACTGTCTGAAAAGCTTGC 783  
DB 23 GTTAAACTGTCTGAAAAGCTTGC 1

Search completed: October 9, 2005, 03:24:42  
Job time : 3173 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2005, 23:39:32 ; Search time 571 Seconds  
(without alignments)  
10159.976 Million cell updates/sec

Title: US-10-829-118-1

Perfect score: 980

Sequence: 1 ctcgcggcggaacacgcctg.....ctgcaccacggtaatgctga 980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	980	100.0	980	6	AAD24180 Human sit
2	980	100.0	2469	6	AAD24181 Human sit
C 3	124.4	12.7	4338	3	AAAS7197 Human sub
C 4	124.4	12.7	4338	6	AAD24182 Human sit
C 5	124.4	12.7	4338	10	ADG32883 Human DNA
C 6	124.4	12.7	4338	13	ADR25092 Breast ca
C 7	64	6.5	114955	2	AAAS3491 Human ade
C 8	60.6	6.2	114955	2	AAAS3491 Human ade
C 9	53	5.4	110000	4	AAI99683_16 Continuation (17 o
C 10	52.6	5.4	4667	12	ADQ22939 Human sof
C 11	51.2	5.2	2000	8	ADA71938 Rice gene
C 12	50.8	5.2	2307	6	ABK32842 DNA encod
C 13	50.8	5.2	4542	10	ADC30607 Human nov
C 14	50.8	5.2	4650	10	ADC32422 Human nov
C 15	50.2	5.1	24000	3	AAAS8551 Human dua
C 16	49.8	5.1	849	6	ABQ29961 Oligonuc
C 17	49.8	5.1	849	6	ABQ29960 Oligonuc
C 18	48.8	5.0	771	6	AAD41042 Human COL
C 19	48.8	5.0	771	6	ABT03863 Human Goo
C 20	48.8	5.0	771	6	ABT03864 Human Goo

C 21	48.8	5.0	771	9	ACD13467 Human Pol
C 22	48.8	5.0	771	9	ACD13466 Human GPB
C 23	48.8	5.0	771	9	ADA97869 Human tum
C 24	48.8	5.0	771	9	ADA97868 Human tum
C 25	48.8	5.0	955	6	AAD41041 Human 955
C 26	48.8	5.0	955	6	ABT03862 Human Goo
C 27	48.8	5.0	955	9	ACD13465 Human pol
C 28	48.8	5.0	955	9	ADA97867 Human tum
C 29	48.8	5.0	40668	6	ABQ88150 Human ost
C 30	48.4	4.9	4225	10	ABV75290 Mouse SCA
C 31	48.2	4.9	1187	12	ACH91930 Human gen
C 32	48	4.9	38847	11	ACN44426 Human gen
C 33	47.6	4.9	989	3	AAA02539 Human col
C 34	47.4	4.8	154746	6	AAD25519 Human her
C 35	47.4	4.8	154746	6	AAD25519 Human her
C 36	46.8	4.8	1535	9	AAD58542 Human thy
C 37	46.8	4.8	1609	13	ADR07763 Full leng
C 38	46.8	4.8	2571	10	ADC63369 Human cdn
C 39	46.8	4.8	2571	10	ABX93989 cDNA enco
C 40	46.8	4.8	2571	12	ADN49355 Human kin
C 41	46.8	4.8	2856	9	ADA38100 Thymidine
C 42	46.8	4.8	2856	10	ADC39135 Novel hum
C 43	46.8	4.8	2856	11	ADM10601 Human thy
C 44	46.8	4.8	2856	12	ADH42558 Novel hum
C 45	46.8	4.8	2893	13	ACN37638 Tumour-as

ALIGNMENTS

RESULT 1

AAD24180

ID AAD24180 standard; DNA; 980 BP.

XX AAD24180;

XX

DT 07-MAY-2002 (first entry)

XX

XX Human site-1 protease promoter partial sequence.

XX

KW Site-1 protease; S1P; promoter; transcription; obesity; type II diabetes;  
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;  
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200200873-A1.  
XX  
PD 03-JAN-2002.  
XX  
PF 19-JUN-2001; 2001WO-SE001386.  
XX  
PR 27-JUN-2000; 2000SE-00002417.  
XX  
(BIOV-) BIOVITRUM AB.  
XX  
PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;  
XX  
DR WPI; 2002-139918/18.  
XX  
PT Human site-1 protease promoter region for identifying agents capable of  
PT inhibiting the promoter activity useful in treating medical conditions  
PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia.  
XX  
Example 1; Page 15; 36pp; English.  
XX  
CC The invention relates to human site-1 protease (S1P) promoter region. The  
CC promoter sequence is useful for identification of compounds that inhibit  
CC transcription of S1P, which in turn results in inhibition of sterol  
CC regulatory element-binding protein (SREBP) pathway. The compound  
CC identified is useful for the treatment of medical conditions related to  
CC obesity, type II diabetes, hypercholesterolaemia, dyslipidaemia,  
CC atherosclerosis and other cardiovascular diseases. The present sequence

CC is a fragment of human SLP promoter  
XX Sequence 980 BP; 175 A; 313 C; 326 G; 166 T; 0 U; 0 Other;  
SQ Query Match 100.0%; Score 980; DB 6; Length 980;  
Best Local Similarity 100.0%; Pred. No. 6.4e-235; Mismatches 0; Indels 0; Gaps 0;  
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTCGCGGGGAAACACACGCTGGGACATCCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
DB 1 CTCGCGGGGAAACACACGCTGGGACATCCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
QY 61 ACTGGGCGGCGGGCCAGGAGGCCACAGCTGGAGCCTCAGCTCCGGCGGACCCAGCGT 120  
DB 61 ACTGGGCGGCGGGCCAGGAGGCCACAGCTGGAGCCTCAGCTCCGGCGGACCCAGCGT 120  
QY 121 GCCCTCTCTGTCGCGCTCCCGGGCTTGGCTGGCGCTCTGGAGCGCTGGGAGCGG 180  
DB 121 GCCCTCTCTGTCGCGCTCCCGGGCTTGGCTGGCGCTCTGGAGCGCTGGGAGCGG 180  
QY 181 GACACGCGGGAGGATGGAAGAAGTGTCTCGGACATTTGGCGCGCGGGGCGCGGTGG 240  
DB 181 GACACGCGGGAGGATGGAAGAAGTGTCTCGGACATTTGGCGCGCGGGGCGCGGTGG 240  
QY 241 CAGGCTGGAAGGAGGAGGCGTGGCCAGGAGCTGCCAGGCGGCGGAGAAACGCGTGGGG 300  
DB 241 CAGGCTGGAAGGAGGAGGCGTGGCCAGGAGCTGCCAGGCGGCGGAGAAACGCGTGGGG 300  
QY 301 AACCTTGTCTCGCTCTGCGGCTCTAGATGATCCCGAAAGGAGCAGCGGCGGAAA 360  
DB 301 AACCTTGTCTCGCTCTGCGGCTCTAGATGATCCCGAAAGGAGCAGCGGCGGAAA 360  
QY 361 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGCGGAGGCTTCCCTGCCCCCGCG 420  
DB 361 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGCGGAGGCTTCCCTGCCCCCGCG 420  
QY 421 GCTGTGTCTATGGAATCTGTGGAGAGCTGCTTCCCGCGGAGACCTTCTGAGAGGGTTC 480  
DB 421 GCTGTGTCTATGGAATCTGTGGAGAGCTGCTTCCCGCGGAGACCTTCTGAGAGGGTTC 480  
QY 481 CAGTCCAGGACCGGCGGCTGGACACCCACCCCGGCGGACCTGCTTGGTGGTGC 540  
DB 481 CAGTCCAGGACCGGCGGCTGGACACCCACCCCGGCGGACCTGCTTGGTGGTGC 540  
QY 541 CCTTAAACCGGCGGTAGCTCTTAAAGATGGGAAAGTGTCCGGTCCGGAACACGCGAAA 600  
DB 541 CCTTAAACCGGCGGTAGCTCTTAAAGATGGGAAAGTGTCCGGTCCGGAACACGCGAAA 600  
QY 601 CCCCATAATCCGCTGCGGACCTCTGTACCCCGGCGGCGGACGAGACTGGGCG 660  
DB 601 CCCCATAATCCGCTGCGGACCTCTGTACCCCGGCGGCGGACGAGACTGGGCG 660  
QY 661 TCCCGAGCGGACGCGCTGCGGGACACCGGTGGCGGAACGAGGACCTTTGTAACT 720  
DB 661 TCCCGAGCGGACGCGCTGCGGGACACCGGTGGCGGAACGAGGACCTTTGTAACT 720  
QY 721 GCCACGTGTTGCTCTTTTGAATAAACAAGATAAATGTGTTAAACTGTCTGAAAGCT 780  
DB 721 GCCACGTGTTGCTCTTTTGAATAAACAAGATAAATGTGTTAAACTGTCTGAAAGCT 780  
QY 781 TGCGGCTTAAAGATGTCTGGGTGAATTAGATGCTAGGATCAGTTGTTTCAATGTAA 840  
DB 781 TGCGGCTTAAAGATGTCTGGGTGAATTAGATGCTAGGATCAGTTGTTTCAATGTAA 840  
QY 841 TGGAACGCGGACCTCCGTACGGGACCTAGCGGACCTGAAGCGCTCTTCAGGTACTGC 900  
DB 841 TGGAACGCGGACCTCCGTACGGGACCTAGCGGACCTGAAGCGCTCTTCAGGTACTGC 900  
QY 901 TGTGTGGCGGTGATGCGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAATTTGACA 960  
DB 901 TGTGTGGCGGTGATGCGCTACAGGCGGATCAGACAGTTTGTCTTCTGGAATTTGACA 960  
QY 961 CTGCACCAACGGTAATGCTGA 980

DB 961 CTGCACCAACGGTAATGCTGA 980  
RESULT 2  
AAD24181  
ID AAD24181 standard; DNA; 2469 BP.  
XX AAD24181;  
XX AC  
XX DT  
XX 07-MAY-2002 (first entry)  
XX Human site-1 protease promoter region.  
XX Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;  
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;  
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.  
XX Homo sapiens.  
XX WO200200873-A1.  
XX PN 03-JAN-2002.  
XX PD  
XX PF 19-JUN-2001; 2001WO-SE001386.  
XX PR 27-JUN-2000; 2000SE-00002417.  
XX PA (BIOV-) BIOVITRUM AB.  
XX PI Abrahamsen L, Ekblom J, Forsgren M, Hoerling J, Johansson P;  
XX WPI; 2002-139918/18.  
XX Human site-1 protease promoter region for identifying agents capable of  
inhibiting the promoter activity useful in treating medical conditions  
such as obesity, diabetes, atherosclerosis and hypercholesterolemia.  
XX Claim 1; Page 15-16; 36pp; English.  
XX The invention relates to human site-1 protease (SLP) promoter region. The  
promoter sequence is useful for identification of compounds that inhibit  
transcription of SLP, which in turn results in inhibition of sterol  
regulatory element-binding protein (SREBP) pathway. The compound  
identified is useful for the treatment of medical conditions related to  
obesity, type II diabetes, hypercholesterolaemia, dyslipidaemia,  
atherosclerosis and other cardiovascular diseases. The present sequence  
is human SLP promoter region  
XX Sequence 2469 BP; 541 A; 704 C; 719 G; 505 T; 0 U; 0 Other;  
SQ Query Match 100.0%; Score 980; DB 6; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 8.3e-235;  
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTCGCGGGGAAACACGCTGGGACATCCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
DB 353 CTCGCGGGGAAACACGCTGGGACATCCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 412  
QY 61 ACTGGGCGGCGGGCCAGGAGGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120  
DB 413 ACTGGGCGGCGGGCCAGGAGGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 472  
QY 121 GCCCTGTCTGTCTCCCGCTCCCGGGCTTGGCTGGCGCTCTGAGCGCGCTGGGACGCG 180  
DB 473 GCCCTGTCTGTCTCCCGCTCCCGGGCTTGGCTGGCGCTCTGAGCGCGCTGGGACGCG 532  
QY 181 GACCAACGCGGAGGATGGAAGAAGTGTCTCGGACATTTGGCGGCGGCGGGGCCCGGTGG 240  
DB 533 GACCAACGCGGAGGATGGAAGAAGTGTCTCGGACATTTGGCGGCGGCGGGGCCCGGTGG 592  
QY 241 CAGGCTGGAAGGAGGCGGCTGGCCAGGAGCTCCAGGCGGCGGAGAACGCGCTGGGGG 300









QY 476 GGGTCCACGTCACGGCCGCGCTCGGACACCCACCCCGCGCGACCTGCGCTG 535  
Db 105261 CGCGCNHNHNSCGGCCCGCGCGCGCGCGCNHNHNSCGGCCCGCGCGCGCGC 105202  
QY 536 GGTGCGCCCTTAACCCCGGGCGGTAGCTCTGTTAAGATGGCGAAGTGTCCGGTCCGGAACACG 595  
Db 105201 CCNHNHNSCGGCCCGCGCGCGCGCCVNNHNHNSCGGCCCGCGCGCGCGCGC 105142  
QY 596 CGAAACCCCAATCCGCGCTGCGCGACCTCTGACCCCGCGGCCCAACGGGACGACGACT 655  
Db 105141 CVGNHNHNSCGGCCCGCGCGCGCGCGCCVGNHNHNSCGGCCCGCGCGCGCGC 105082  
QY 656 GGGCTCCGCGCGCGCGCGCG 677  
Db 105081 GCCVGGCNHNHNSCGGCCCG 105060

RESULT 8

AA53491  
ID AAX53491 standard; DNA; 114955 BP.  
XX  
AC AAX53491;  
XX  
XX  
DT 05-JUL-1999 (first entry)  
XX  
DE Human adenosine A1 receptor antisense oligonucleotide fragment.  
XX  
KW Antisense oligonucleotide; multiple target; antisense treatment;  
KW impaired respiration; inflammation; lung disease;  
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
KW acute asthma; allergy; asthma; impeded respiration;  
KW respiratory distress syndrome; pain; cystic fibrosis;  
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
KW prostate cancer; ss.  
XX  
OS Synthetic.  
XX  
XX  
PN WO9913886-A1.  
XX  
PD 25-MAR-1999.  
XX  
PF 17-SEP-1998; 98WO-US019419.  
XX  
PR 17-SEP-1997; 97US-0059160P.  
PR 09-JUN-1998; 98US-00093972.  
XX  
XX (UYEC-) UNIV EAST CAROLINA.  
PA  
PI Nyce JW;  
XX  
XX WPI; 1999-229400/19.  
XX  
XX New antisense oligonucleotides used in treatment of, e.g. pulmonary  
XX vasoconstriction.  
XX  
XX Disclosure; Page 37; 120pp; English.  
XX

CC The specification describes antisense oligonucleotides (AA52869-X55271)  
CC directed against at least 2 mRNAs selected from target genes, coding and  
CC non-coding regions of RNAs corresponding to target genes, gene initiation  
CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-  
CC end and the juxta-section between coding and non-coding regions and all  
CC segments of RNAs encoding proteins associated with one or more diseases,  
CC conditions or mixtures. The antisense oligonucleotides may be derived  
CC from sequences AA55272-74. These multiple target oligonucleotides  
CC (specifically AA55180-271) can be used for the antisense treatment of  
CC diseases and conditions. Typical diseases and conditions are those  
CC associated with impaired respiration and inflammation, including lung  
CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,

CC acute asthma, allergies, asthma, impeded respiration, respiratory  
CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,  
CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary  
CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.  
CC colon cancer, breast cancer, lung cancer, pancreatic cancer,  
CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as  
CC well as all types of cancers which may metastasize or have metastasized  
CC to the lungs, including breast and prostate cancer  
XX

SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;  
Query Match 6.2%; Score 60.6; DB 2; Length 114955;  
Best Local Similarity 32.1%; Pred. NO. 0.0002;  
Matches 210; Conservative 62; Mismatches 377; Indels 6; Gaps 1;

QY 32 CGGGGCTGTTTACTCCCAACTCTCGGAGACTGGGGGCGCGGCCAGGCGCCACAGC 91  
Db 104692 CBGGCCBGCGCGCGCSNNNDNNCCGCGCGCGCGCGCSNNNDNNCCGCGCGCGC 104751  
QY 92 TGGGAGCTCTAGCTCCGCGCACCCAGCTGCTCTGTCTCCGCGCTCCCGGGGCTTGC 151  
Db 104752 GCSNNNDNNCCGCGCGCGCGCSNNNDNNCCGCGCGCGCGCSNNNDNNCCGCGCGC 104811  
QY 152 GTGCGCGCTCTGGACGCCGTGGGACGCGGACACCGCGCGGAGGATGACGAGGTGCTC 211  
Db 104812 CBGGGCSNNNDNNCCGCGCGCGCSNNNDNNCCGCGCGCGCSNNNDNNCCGCGCGCBS 104871  
QY 212 GCGACATTTGCG 271  
Db 104872 NNNNDNNCCGCGCGCGCSNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104931  
QY 272 GCTGCCAGCGCGCGAGAACGCGCTGCGGGGAAACCTTGTGCTGCTGCGCGCTCTAG 331  
Db 104932 GCGGCG 104991  
QY 332 GATCCCCGAAAGAGAGCAGCGCGCGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 391  
Db 104992 GGGCGCSNNNDNNBGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105051  
QY 392 TGCTTGGGCGAGGCTCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 445  
Db 105052 CGCGCGCGCGCGCGCSNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105111  
QY 446 GCTGCTTCCCG 505  
Db 105112 GCG 105171  
QY 506 CACCCCAACCCCG 565  
Db 105172 GCG 105231  
QY 566 AAGATGGGGAAGTGTCCGCTCCGGAACACGCGGAACCCCAATCCGCGCTGCGCGACCTC 625  
Db 105232 GCG 105291  
QY 626 CTGACCCCG 680  
Db 105292 CGCGCGCGCGCGCSNNNDNNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105346

RESULT 9

AA199683\_16/c  
Continuation (17 of 44) of AA199683 from base 1600001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683  
WP Fragment Name Begin End  
WP AA199683\_00 1 110000  
WP AA199683\_01 100001 210000  
WP AA199683\_02 200001 310000  
WP AA199683\_03 300001 410000  
WP AA199683\_04 400001 510000  
WP AA199683\_05 500001 610000  
WP AA199683\_06 600001 710000  
WP AA199683\_07 700001 810000













QY 429 TCATGACTCGTGGAGAGCTCGCTTCCGCGCGGACCCCTTCCTGACAGGGTCCACGTCCA 488  
 Db |||||  
 4372 ACCTGGCGGGGGCGGCTCCGCGCGCGCTCCGCGCGCGCCCTCGCGCGGGCGCGGCC 4431  
 QY 489 GGCACCGCGGCTCCGAC 506  
 Db |||||  
 4432 TCGCGCGCGGCTCCGCGC 4449

## RESULT 15

AAA88551/c

ID AAA88551 standard; DNA; 24000 BP.

XX

AC AAA88551;

XX

DT 22-JAN-2001 (first entry)

XX

DE Human dual-specificity phosphatase-1 (DSP-1) gene.

XX

KW DSP-1; dual-specificity phosphatase-1; human; cell proliferation;  
 cell differentiation; cell survival; cell cycle; dephosphorylation;  
 KW signal transduction; MAP-kinase; cancer; graft versus host disease;  
 KW allergy; autoimmune disease; metabolic disease; therapy; chromosome 17;  
 KW db.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT exon

42..109

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

XX

The present sequence is that of the human dual-specificity phosphatase-1 (DSP-1) gene on chromosome 17. The gene was identified in genomic sequences obtained from an expressed sequence tag database screened with a conserved motif (see AAB19604) of known DSPs. DSP-1 dephosphorylates both phosphothreonine/serine and phosphotyrosine residues in DSP-1 substrates such as activated mitogen-activated protein kinase (MAP-kinase). DSP-1 has sequence homology to other MAP-kinase phosphatases. It is expressed at high levels in the human heart, testis and liver, and at lower levels in other tissues. Methods are provided for recombinant production of DSP-1 polypeptides, and for using DSP-1 polypeptides, antibodies and polynucleotides to detect DSP-1 expression, to screen for agents that modulate DSP-1 activity e.g. within a combinatorial library, and for using such agents to modulate cell proliferation, cell differentiation or cell survival, through modulation of pattern of gene expression, apoptosis or cell cycle. In particular, the cell displays contact inhibition of cell growth, anchorage-dependent growth or an altered intercellular adhesion property, or is a cell present in a patient afflicted with a disorder associated with DSP-1 activity, such as cancer, graft-versus host disease, autoimmune disease, allergy, metabolic disease, abnormal cell growth, abnormal cell proliferation and abnormal cell cycle

XX

SQ Sequence 24000 BP; 6038 A; 5505 C; 5602 G; 6855 T; 0 U; 0 Other;

## Query Match

Best Local Similarity 5.1%; Score 50.2; DB 3; Length 24000;

Matches 171; Conservative 0; Mismatches 183; Indels 1; Gaps 1;

QY 186 CGCCGGGAGGATGGACGAAGGTGCTCGCACATTTTCGCGCGGGCGGGCGGAGGG 245  
 Db |||||  
 435 CGCGCGCGGAGGCTGCTGGAGCCCGGGGCACAGGGACAGGGTCGCGGGGCGCGCGCG 376  
 QY 246 TGGAGCGGAGGGCGTGGCCAGCGAGCTGCCAGCGCGGAGAACCGCTGGGGGAACCC 305  
 Db |||||  
 375 CGCGCGGGGAGCGCGAGGGGGCGCACGCTCAGGGCCCGCGAGGGTGCCTGCCGAAGGTC 316  
 QY 306 TTGGTCCGCTCTGCGCGTCTAGGATCCCCGAAAAGGAGCACGGGGCGCGAAAGCGGC 365  
 Db |||||  
 315 TCGGGCTCTCTCTC-CGCGGGCGCCCGCAAGAGTCAAGGACCGCCAGGGAGCAGGTGGG 257  
 QY 366 CAGGCTGGGCCAGGATCTAGAAAGACTGCTCGCGCAGCGCTCCCTGCCCGCGGGCGCTG 425  
 Db |||||  
 256 GCGCGCGGCACAAAGTTTCCAGAGGCGGGTTCGCGCGGGCTCTGGGTCTCCCTGGGCC 197  
 QY 426 CTGTATGACTCGTGGAGAGCTCGCTTCCCGCGGAGACCTTCTTCGAGGGGTCCACGT 485  
 Db |||||  
 196 AGGGACCCGAGCGGAGTCCCGCCCTCCCGCCCCACCCCAACACCTTCCCCCACAACCC 137  
 QY 486 CCAGGCACCGCGGCTCGGACACACCCACCGCGCGGCGACCTGCTCCCTGGGTGC 540  
 Db |||||  
 136 CCACCCCGGGCTCGGGGTCTCCCTACCGCGGGCGCGCGGGCGGGTTCG 82

Search completed: October 9, 2005, 02:31:41

Job time : 575 secs

New isolated polypeptide having the sequence of dual-specificity phosphatase-1 (DSP-1) is useful for treating a patient with a disorder associated with DSP-1 activity e.g. cancer and autoimmune diseases.  
 Example 1; Fig 3A-J; 74pp; English.

***This Page Blank (uspto)***

***This Page Blank (uspto)***

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2005, 22:04:07 ; Search time 4298 Seconds  
(without alignments)  
11048.421 Million cell updates/sec

Title: US-10-829-118-1  
Perfect score: 980  
Sequence: 1 ctcgcggcggaacgcctg.....ctgcaccacggaatgctga 980

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pri.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	980	100.0	980	6 AR527527	AR527527 Sequence
2	980	100.0	2469	6 AR527528	AR527528 Sequence
3	964.8	98.4	212596	9 AC040169	AC040169 Homo sapi
4	964.8	98.4	216522	2 AC139664	AC139664 Homo sapi
5	921.4	53.2	532	9 BC052334	BC052334 Homo sapi
6	126	12.9	4338	6 CQ729132	CQ729132 Sequence
7	124.4	12.7	4338	6 BD244989	BD244989 Mammalian
8	124.4	12.7	4338	6 AR261390	AR261390 Sequence
9	124.4	12.7	4338	6 AR527529	AR527529 Sequence
10	124.4	12.7	4338	6 AX024613	AX024613 Sequence
11	124.4	12.7	4338	9 HUMK1AAL	D42053 Homo sapien
12	90.4	9.2	3026	9 BC026330	BC026330 Homo sapi
13	64.4	6.6	1393	11 PM11H12G	AL684264 Penicilli
14	64.2	6.6	125020	9 AF429315	AF429315 Homo sapi
15	63.8	6.5	300695	2 AC079431	AC079431 Mus muscu
16	63.6	6.5	219952	2 AC084804	AC084804 Mus muscu
17	62.2	6.3	65351	2 AC139773	AC139773 Homo sapi
18	60.6	6.2	155337	2 AC116408	AC116408 Mus muscu
19	59.8	6.1	1065	11 PM2B12B	AL684695 Penicilli

59.6	6.1	1200	11	PM12A10G	AL684284 Penicilli
59.4	6.1	1200	11	PM12A10G	AL684284 Penicilli
59.4	6.1	159028	2	AC142003	AC142003 Rattus no
59.4	6.1	208238	2	CR385024	CR385024 Danio rer
59.2	6.0	125020	9	AF429315	AF429315 Homo sapi
59	6.0	132772	9	AL590105	AL590105 Human DNA
59	6.0	239130	2	AC079420	AC079420 Mus muscu
59	6.0	245210	2	AC137771	AC137771 Homo sapi
58.8	6.0	53552	2	AC100796	AC100796 Homo sapi
58.6	6.0	1007	11	PM3H11G	AL685749 Penicilli
58.4	6.0	1007	11	PM3H11G	AL685749 Penicilli
58.2	5.9	972	11	PM12A12G	AL684288 Penicilli
58.2	5.9	136753	2	AC149968	AC149968 Strongylo
58.2	5.9	159028	2	AC142003	AC142003 Rattus no
57.4	5.9	85434	2	AC066610	AC066610 Homo sapi
57.2	5.8	931	11	PM4A12G	AL685775 Penicilli
57	5.8	136753	2	AC149968	AC149968 Strongylo
56.8	5.8	155337	2	AC116408	AC116408 Mus muscu
56.2	5.7	931	11	PM4A12G	AL685775 Penicilli
56	5.7	1393	11	PM11H12G	AL684264 Penicilli
56	5.7	277058	2	AC016159	AC016159 Drosophi
55.8	5.7	47958	2	AC091104	AC091104 Homo sapi
55.8	5.7	217412	2	AC024400	AC024400 Homo sapi
55.6	5.7	991	11	PM12H12B	AL684455 Penicilli
55.6	5.7	1065	11	PM2B12B	AL684695 Penicilli
55.2	5.6	136222	9	AC004099	AC004099 Homo sapi

ALIGNMENTS

RESULT 1  
AR527527  
LOCUS AR527527 980 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 1 from patent US 6723553.  
ACCESSION AR527527  
VERSION AR527527.1 GI:53914632  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 980)  
AUTHORS Abrahmsen, L., Ekblom, J., Forsgren, M., Horling, J. and Johansson, P.  
TITLE Promoter sequences  
PROMOTER SEQUENCES  
JOURNAL Patent: US 6723553-A 1 20-APR-2004;  
FEATURES  
source  
1..980  
/organism="unknown"  
/mol\_type="genomic DNA"

Query Match	100.0%	Score	980	DB	6	Length	980
Best Local Similarity	100.0%	Pred. No.	2.1e-196	Mismatches	0	Indels	0
Matches	980	Conservative	0	0	0	Gaps	0
QY	1	CTCCGGCGGGAACACACGCCTGGGCACCTCCATTCGGGGCTGTTTACTCCCACTCTCGCGAG	60				
Db	1	CTCCGGCGGGAACACACGCCTGGGCACCTCCATTCGGGGCTGTTTACTCCCACTCTCGCGAG	60				
QY	61	ACTGGGCGCGCGGGCCAGCGAGCCACAGCTGGAGCCTCAGCTCCGGCGAGCCCGAGCGT	120				
Db	61	ACTGGGCGCGCGGGCCAGCGAGCCACAGCTGGAGCCTCAGCTCCGGCGAGCCCGAGCGT	120				
QY	121	GCCTGTCTGTCTCCCGCTCCCGGGCTTGGCGCGCTCTGGAGCGCGTGGGCGAGCGG	180				
Db	121	GCCTGTCTGTCTCCCGCTCCCGGGCTTGGCGCGCTCTGGAGCGCGTGGGCGAGCGG	180				
QY	181	GACACCGCGGAGGATGACGAGTGTCCGGACATTTCCGGCGCGCGGGCGCGGTGG	240				
Db	181	GACACCGCGGAGGATGACGAGTGTCCGGACATTTCCGGCGCGCGGGCGCGGTGG	240				
QY	241	CAGGGTGAAGGAGGCGGGCGGTGGCCAGCGAGCTGCAGCGCGCGAGAACGCGCTGGGG	300				
Db	241	CAGGGTGAAGGAGGCGGGCGGTGGCCAGCGAGCTGCAGCGCGCGAGAACGCGCTGGGG	300				

Db	241	CAGGGTGAAGCGGAGGGGCGTGGCCAGCGAGCTGCCAGCGCGGCGAGAAACGGCGCTGGGG	300
Qy	301	AACCTTGTGTCGGCTCTGGCGCTCTAGGATCCCGAAAGAGGACACGGGCGCGAAA	360
Db	301	AACCTTGTGTCGGCTCTGGCGCTCTAGGATCCCGAAAGAGGACACGGGCGCGAAA	360
Qy	361	GCGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCCCGGG	420
Db	361	GCGGCAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCCCGGG	420
Qy	421	GCTGCTGTCAATGAGTCTGGGAGAGCTGCTTCCCGCGGACCTTCTCTCAGGGTTC	480
Db	421	GCTGCTGTCAATGAGTCTGGGAGAGCTGCTTCCCGCGGACCTTCTCTCAGGGTTC	480
Qy	481	CAGGTCCAGGACACCGCGCTCGGACACCCCAACCCCGCGGCGACCTGCCCTGGGTGC	540
Db	481	CAGGTCCAGGACACCGCGCTCGGACACCCCAACCCCGCGGCGACCTGCCCTGGGTGC	540
Qy	541	CCCTTAAACCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGCTCCGGAACACCGGAAA	600
Db	541	CCCTTAAACCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGCTCCGGAACACCGGAAA	600
Qy	601	CCCCAAATCCCGCTGCCGACCTCTGACCCCGCGGCCCAACCGGACGACAGCTGGGC	660
Db	601	CCCCAAATCCCGCTGCCGACCTCTGACCCCGCGGCCCAACCGGACGACAGCTGGGC	660
Qy	661	TCCGACGCGCAGCGCGCTGCCGGGACACCGGTGGCTGGGAAACGAGGACCTTTGTAAAC	720
Db	661	TCCGACGCGCAGCGCGCTGCCGGGACACCGGTGGCTGGGAAACGAGGACCTTTGTAAAC	720
Qy	721	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	780
Db	721	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	780
Qy	781	TGCGGCTAAAGATGTCTGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	840
Db	781	TGCGGCTAAAGATGTCTGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	840
Qy	841	TGACACAGCCCGGACTCCGTACGGCAGTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	900
Db	841	TGACACAGCCCGGACTCCGTACGGCAGTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	900
Qy	901	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCGAACTTGACA	960
Db	901	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCGAACTTGACA	960
Qy	961	CTGCACACGGTAATGCTGA	980
Db	961	CTGCACACGGTAATGCTGA	980

RESULT 2  
AR257528  
LOCUS AR257528 2469 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 2 from patent US 6723553.  
ACCESSION AR257528  
VERSION AR257528.1 GI:53914633  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 2469)  
AUTHORS Abrahamson, L., Ekblom, J., Forsgren, M., Horling, J. and Johansson, P.  
TITLE Promoter sequences  
JOURNAL Patent: US 6723553-A 2 20-APR-2004;  
FEATURES  
    1..2469  
        Location/Qualifiers  
            /mol\_type="genomic DNA"  
ORIGIN

Query Match 100.0%; Score 980; DB 6; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 2e-196;

Matches	980;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	CTCCGCGCGGAAACACGCGCTGGGCACTCATTTCCGGGCTGTTTACTCCCAACTCTCCGAG	60						
Db	353	CTCCGCGCGGAAACACGCGCTGGGCACTCATTTCCGGGCTGTTTACTCCCAACTCTCCGAG	412						
Qy	61	ACTGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	120						
Db	413	ACTGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	472						
Qy	121	GCCCTGTCTGTCTCCCGCGCTCCCGGCGCTTGGGTGGCGCTCTTGACGCGCGTGGGAGCGG	180						
Db	473	GCCCTGTCTGTCTCCCGCGCTCCCGGCGCTTGGGTGGCGCTCTTGACGCGCGTGGGAGCGG	532						
Qy	181	GACCAACCGCGGAGGATGGAAGAGTCTCGCGACATTTGGGCGGCGGCGGCGGCGGCGGCGG	240						
Db	533	GACCAACCGCGGAGGATGGAAGAGTCTCGCGACATTTGGGCGGCGGCGGCGGCGGCGGCGG	592						
Qy	241	CAGGCTGGAAGCGGAGGGGCGTGGCCAGCGAGCTGCCAGGGCGGCGGAGAACCGCTGGGGG	300						
Db	593	CAGGCTGGAAGCGGAGGGGCGTGGCCAGCGAGCTGCCAGGGCGGCGGAGAACCGCTGGGGG	652						
Qy	301	AACCTTGTCTCGCTCTCGCGCTCTAGGATCCCGGAAAGAGGACACGGGCGCGGAAA	360						
Db	653	AACCTTGTCTCGCTCTCGCGCTCTAGGATCCCGGAAAGAGGACACGGGCGCGGAAA	712						
Qy	361	GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	420						
Db	713	GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	772						
Qy	421	GCTGCTGTCTGATGAGTCTGGGAGAGCTGCTTCCCGCGCGGACCTTCTCGAGGGTTC	480						
Db	773	GCTGCTGTCTGATGAGTCTGGGAGAGCTGCTTCCCGCGCGGACCTTCTCGAGGGTTC	832						
Qy	481	CACGTCCAGGACACCGCGGCTCGGACACCCACACCCCGCGCGGCGGCGGCGGCGGCGG	540						
Db	833	CACGTCCAGGACACCGCGGCTCGGACACCCACACCCCGCGCGGCGGCGGCGGCGGCGG	892						
Qy	541	CCCTTAAACCGGCGGCGTGTCTGTTAAGATGGCGAAGTGTCCGCTCCGGAACACCGGAAA	600						
Db	893	CCCTTAAACCGGCGGCGTGTCTGTTAAGATGGCGAAGTGTCCGCTCCGGAACACCGGAAA	952						
Qy	601	CCCCAAATCCCGCTGCCGACCTCTGACCCCGCGGCCCAACCGGACGACAGCTGGGC	660						
Db	953	CCCCAAATCCCGCTGCCGACCTCTGACCCCGCGGCCCAACCGGACGACAGCTGGGC	1012						
Qy	661	TCCGACGCGCAGCGCTGCCGGGACACCGGTGGGTGGGAAACGAGGAGGACCTTTGTAAAC	720						
Db	1013	TCCGACGCGCAGCGCTGCCGGGACACCGGTGGGTGGGAAACGAGGAGGACCTTTGTAAAC	1072						
Qy	721	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	780						
Db	1073	GCCAGCTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAAGCT	1132						
Qy	781	TGCGGCTAAAGATGTCTGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	840						
Db	1133	TGCGGCTAAAGATGTCTGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGTAAA	1192						
Qy	841	TGACACAGCCCGGACTCCGTACGGCAGTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	900						
Db	1193	TGACACAGCCCGGACTCCGTACGGCAGTAGCAGGGGACTGAAAGCGCTCTTCAGGTACTGC	1252						
Qy	901	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCGAACTTGACA	960						
Db	1253	TGTTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTTCGAACTTGACA	1312						
Qy	961	CTGCACACGGTAATGCTGA	980						
Db	1313	CTGCACACGGTAATGCTGA	1332						

RESULT 3  
AC040169

```
LOCUS AC040169 212596 bp DNA linear PRI 17-DEC-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-505K9, complete sequence.
ACCESSION AC040169
VERSION AC040169.7 GI:271511363
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212596)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 212596)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 212596)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 17, 2002 this sequence version replaced gi:18139274.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.4.
NOTE: Shatter libraries failed to resolve dinucleotide repeat
region from 8264 to 8629. Unsure number of repeat copies 8264 to
8629. Forced join at 8417.
FEATURES
    Location/Qualifiers
        1..212596
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosomes="16"
            /clone="RP11-505K9"
            8264..8629
            /note="NOTE: Shatter libraries failed to resolve
            dinucleotide repeat region from 8264 to 8629. Unsure
            number of repeat copies 8264 to 8629. Forced join at
            8417."
        misc_feature
            /note="NOTE: Shatter libraries failed to resolve
            dinucleotide repeat region from 8264 to 8629. Unsure
            number of repeat copies 8264 to 8629. Forced join at
            8417."
    ORIGIN
        Query Match 98.4%; Score 964.8; DB 9; Length 212596;
        Best Local Similarity 99.7%; Pred. NO. 2.8e-193;
        Matches 977; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
        QY 1 CTCGCGGGAACAGCGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTGCGGAG 60
        DB 173307 CTCGCGGGAACAGCGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTGCGGAG 173366
        QY 61 ACTGGGCGGCTGGGCGAGGAGCCGACAGCTGGGAGCTTCAGCTCCGCGGACCCAGCGT 120
        DB 173367 ACTGGGCGGACCGGGCGAGGAGCCGACAGCTGGGAGCTCAGCTCCGCGGACCCAGCGT 173426
        QY 121 GCCCTGTCTGTCCTCCGGCTCCCGGGCTTGCCTGGCGCTCTGGACGCGCTGGGAGCGG 180
        DB 173427 GCCCTGTCTGTCCTCCGGCTCCCGGGCTTGCCTGGCGCTCTGGACGCGCTGGGAGCGG 173486
        QY 181 GACCACGCGGAGGATGGAAGAAGTCTCCGACATTTGCGGCGGCGGGGGCGGCTGG 240
        DB 173487 GACCACGCGGAGGATGGAAGAAGTCTCCGACATTTGCGGCGGCGGGGGCGGCTGG 173545
        QY 241 CAGGCTGGAAGCGGAGGCGTGGCGAGCGAGCTGCCAGGCGGCGAGAACCGCTGGGGG 300
        DB 173546 CAGGCTGGAAGCGGAGGCGTGGCGAGCGAGCTGCCAGGCGGCGAGAACCGCTGGGGG 173605
```

```
QY 301 AACCCCTTGTTCGGCTCTGCGCGTCTCTAGGATCCCGAAAGAGGACGCGGCGCGAAA 360
DB 173606 AACCCCTTGTTCGGCTCTGCGCGTCTCTAGGATCCCGAAAGAGGACGCGGCGCGAAA 173665
QY 361 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGCGGAGGCTCCCTGCCCCCGCGG 420
DB 173666 GCGGCGAGGCTGGGCGAGGATCTAGAAAGACTGCTTGGCGGAGGCTCCCTGCCCCCGCGG 173725
QY 421 GCCTCTGTCTGAGGACTCGTGAGAGCTCGTCTCCCGCGGAGCCCTCTCTGAGGGGTC 480
DB 173726 GCCTCTGTCTGAGGACTCGTGAGAGCTCGTCTCCCGCGGAGCCCTCTCTGAGGGGTC 173785
QY 481 CACGTCACAGGCAACCGGCGGCTCGGACACCCCAACCCCGCGGCGGACCTGCGCTGGGTGC 540
DB 173786 CAGGTCACAGGCAACCGGCGGCTCGGACACCCCAACCCCGCGGCGGACCTGCGCTGGGTGC 173845
QY 541 CCCTTAAACCCCGGCGGTAGCTGTTAAGATGCGGAAGTGTCCGGTCCGGAACACGCGAAA 600
DB 173846 CCCTTAAACCCCGGCGGTAGCTGTTAAGATGCGGAAGTGTCCGGTCCGGAACACGCGAAA 173905
QY 601 CCCCAAAATCCCGCTGCGGACCTCTGACCCCGCGGCGGACCGGACGACACTGGGCC 660
DB 173906 CCCCAAAATCCCGCTGCGGACCTCTGACCCCGCGGCGGACCGGACGACACTGGGCC 173965
QY 661 TCCCGACGCGGACGCGCTGCGGCGGACACCGGTGGGTGCGAAACGAGGACCTTTCTTAAC 720
DB 173966 TCCCGACGCGGACGCGCTGCGGCGGACACCGGTGGGTGCGAAACGAGGACCTTTCTTAAC 174025
QY 721 GCCACGTGTTGCTCTTTTGTGAAACAAAGAAATAAATGTGTTAAACTCTCTGAAAGCT 780
DB 174026 GCCACGTGTTGCTCTTTTGTGAAACAAAGAAATAAATGTGTTAAACTCTCTGAAAGCT 174085
QY 781 TCCCGCTTAAAGATGTCGCGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGATAA 840
DB 174086 TCCCGCTTAAAGATGTCGCGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGATAA 174145
QY 841 TGGACACGCGCGGACTCCGTCAGCGCACTAGCAGGAGGACTGAAAGCGCTCTTCAGGTACTGC 900
DB 174146 TGGACACGCGCGGACTCCGTCAGCGCACTAGCAGGAGGACTGAAAGCGCTCTTCAGGTACTGC 174205
QY 901 TGGTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTCTGGAACCTTGACA 960
DB 174206 TGGTGGGCGGTGATCGCTACAGGCGGATCAGACAGTTTGTCTCTGGAACCTTGACA 174265
QY 961 CTGCACACAGGTAATGCTGA 980
DB 174266 CTGCACACAGGTAATGCTGA 174285
```

## RESULT 4

AC139664/c

LOCUS

DEFINITION

AC139664

Homo sapiens clone RP11-812D13, WORKING DRAFT SEQUENCE, 9 unordered

pieces.

ACCESSION

AC139664

VERSION

AC139664.5

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 212596)

Munz, D.M., Adams, C., Adio-Oduola, B., Ali-oman, F.R., Allen, C.,

Albrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Davila, M.L., Davis, C., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
 Homai, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovich, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, J., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
 Sodergren, B., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 216522)  
 Worley K.C.

Direct Submission  
 Submitted (08-FEB-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 216522)  
 Worley K.C.

Direct Submission  
 Submitted (27-MAR-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Mar 27, 2003 this sequence version replaced gi:28467047.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Drafting Center Code: BCM  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HETI  
 Center clone name: RP11-812D13  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 215023 bases at least Q40  
 Consensus quality: 215949 bases at least Q30  
 Consensus quality: 216649 bases at least Q20  
 Estimated insert size: 210467; sum-of-contigs estimation  
 Quality coverage: 10x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 consists of 9 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will

\* be preserved.

1 1056: contig of 1056 bp in length  
 1057 1156: gap of unknown length  
 1157 2298: contig of 1142 bp in length  
 2299 2398: gap of unknown length  
 2399 3437: contig of 1039 bp in length  
 3438 3537: gap of unknown length  
 3538 5421: contig of 1884 bp in length  
 5422 5521: gap of unknown length  
 5522 6548: contig of 1027 bp in length  
 6549 7859: contig of 1211 bp in length  
 7860 7959: gap of unknown length  
 7960 21513: contig of 13554 bp in length  
 21514 21613: gap of unknown length  
 21614 40787: contig of 19174 bp in length  
 40788 40887: gap of unknown length  
 40888 216522: contig of 175635 bp in length.

Location/Qualifiers  
 1. 216522  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="RP11-812D13"

FEATURES  
 source  
 ORIGIN

Query Match 98.4%; Score 964.8; DB 2; Length 216522;  
 Best Local Similarity 99.7%; Pred. No. 2.8e-193;  
 Matches 977; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CTCGCGCGGACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60  
 DB 19428 CTCGCGCGGACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 19369

QY 61 ACTGGCGCGCGGCGGACGAGGCGCCACAGCTGGAGGCTTCAGTCCCGCAGACCCAGCGT 120  
 DB 19368 ACTGGCGCGCGGCGGACGAGGCGCCACAGCTGGAGGCTTCAGTCCCGCAGACCCAGCGT 19309

QY 121 GCCCTGTCTGTCTCCCGCTCCCGGGCTTGGGTGGCGCTCTGAGACGCCGTGGGAGCGG 180  
 DB 19308 GCCCTGTCTGTCTCCCGCTCCCGGGCTTGGGTGGCGCTCTGAGACGCCGTGGGAGCGG 19249

QY 181 GACACGCGCGGAGGATGAGCAAGGTGCTCGCGACATTTGGCGGGGGGGGGCGGCTGG 240  
 DB 19248 GACACGCGCGGAGGATGAGCAAGGT-CTCGGACATTTGGCGGGGGGGGGCGGCTGG 19190

QY 241 CAGGCTGGAAGCGGAGGCGGTGGCGAGCTGCCAGGCGCGGAGAACGCGCTGGGGG 300  
 DB 19189 CAGGCTGGAAGCGGAGGCGGTGGCGAGCTGCCAGGCGCGGAGAACGCGCTGGGGG 19130

QY 301 AACCTTGTGTCTCGCTCTGCGCGTCTTAGGATCCCGGAAAGAGGACACGGGCGGAAA 360  
 DB 19129 AACCTTGTGTCTCGCTCTGCGCGTCTTAGGATCCCGGAAAGAGGACACGGGCGGAAA 19070

QY 361 GCGCGCAGGCTGGCGGAGGATGAGCAAGGTGCTTCGCGGAGGCTTCCTGCGCCCGCGG 420  
 DB 19069 GCGCGCAGGCTGGCGGAGGATGAGCAAGGTGCTTCGCGGAGGCTTCCTGCGCCCGCGG 19010

QY 421 GCCTGTGTATGGACTCGTGAGAGCTCGCTTCCCGCGGAGACCTTCTCTGAGGGGTC 480  
 DB 19009 GCCTGTGTATGGACTCGTGAGAGCTCGCTTCCCGCGGAGACCTTCTCTGAGGGGTC 18950

QY 481 CAGCTCCAGGACACGGCGGCTCGGACACCCACACCCCGCGGCGACCTTGCCTTGGGTGC 540  
 DB 18949 CAGCTCCAGGACACGGCGGCTCGGACACCCACACCCCGCGGCGACCTTGCCTTGGGTGC 18890

QY 541 CCCTTAAACCGGCGGTAGCTGTTAAGATGCGGAGTCTCCGTCGCGGAAACCGGAAA 600  
 DB 18889 CCCTTAAACCGGCGGTAGCTGTTAAGATGCGGAGTCTCCGTCGCGGAAACCGGAAA 18830

QY 601 CCCCATAATCCCGCTGCGCGACTCTCTGACCCCGCGGCCCGGACGACAGACTCGGGCC 660  
 DB 18829 CCCCATAATCCCGCTGCGCGACTCTCTGACCCCGCGGCCCGGACGACAGACTCGGGCC 18770

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL

COMMENT

QY 661 TCCCGACGCGCAGCGCGCTGCGGGACACCGGTCGCGTGCAGAAACGGAGGACCTTTCTTAAC 720  
Db 18769 TCCCGACGCGCAGCGCGCTGCGGGACACCGGTCGCGTGCAGAAACGGAGGACCTTTCTTAAC 18710  
QY 721 GCCACGTGTTGCTCTTTTGAAGAAACAGATAAATGTTTAAACTGCTCTGAAAGCT 780  
Db 18709 GCCACGTGTTGCTCTTTTGAAGAAACAGATAAATGTTTAAACTGCTCTGAAAGCT 18650  
QY 781 TGCCCGCTAAAGACGTCTGGGTGACTTAGATCTAGGATCAGTTGTTTCAATGTA 840  
Db 18649 TGCCCGCTAAAGACGTCTGGGTGACTTAGATCTAGGATCAGTTGTTTCAATGTA 18590  
QY 841 TGACACGCGCGACCTCCGTACGGCACTAGCAGGGGACCTGAAACGCTCTTCAGGTACTGC 900  
Db 18589 TGACACGCGCGACCTCCGTACGGCACTAGCAGGGGACCTGAAACGCTCTTCAGGTACTGC 18530  
QY 901 TGGTGGCGCGTGTGCGGTACAGCGCGATCAGACAGTTTGTCTCTGGAACCTTGACA 960  
Db 18529 TGGTGGCGCGTGTGCGGTACAGCGCGATCAGACAGTTTGTCTCTGGAACCTTGACA 18470  
QY 961 CTGCACACCGTAATGCTGA 980  
Db 18469 CTGCACACCGTAATGCTGA 18450

BC052334 532 bp mRNA linear PRI 05-MAY-2003  
Homo sapiens cDNA clone IMAGE:6176570, partial cds.  
BC052334  
BC052334.1 GI:30354085

ORGANISM  
Homo sapiens (human)

REFERENCE  
AUTHORS  
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McSwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,  
Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932

2 (bases 1 to 532)  
Straussberg, R.

Direct Submission  
Submitted (02-MAY-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>

Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
Teurisson, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 115 Row: n Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis, Genomescan gene prediction.

## FEATURES

Location/Qualifiers  
source

1..532  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6176570"  
/tissue\_type="uterus, leiomyosarcoma"  
/clone\_lib="NIH\_MGC\_71"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 53.2%; Score 521.4; DB 9; Length 532;  
Best Local Similarity 99.8%; Pred. No. 1.3e-99;  
Matches 522; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 272 GCTGCCAGGCGCGAGAACGCGCTGGGGAAACCTTTGGTCCGCTCTGCGGCTCGCTCTAG 331  
Db 1 GCTGCCAGGCGCGAGAACGCGCTGGGGAAACCTTTGGTCCGCTCTGCGGCTCGCTCTAG 60  
QY 332 GATCCCCGAAAAAGGAGACACGGCGCGGAAACCGGCCAGGCTGGGCCAGGATCTAGAAAGAC 391  
Db 61 GATCCCCGAAAAAGGAGACACGGCGCGGAAACCGGCCAGGCTGGGCCAGGATCTAGAAAGAC 120  
QY 392 TGCCTGGCGCAGGCTCCCTGCCCCCGCGGCTCTGCTCATGACTCTGCGAGAGCTCGC 451  
Db 121 TGCCTGGCGCAGGCTCCCTGCCCCCGCGGCTCTGCTCATGACTCTGCGAGAGCTCGC 180  
QY 452 TTCCCGCGCGGACCCCTTCTCTGAGGGGTCCAGTCCAGGCACCGCGGCTCGGACACCCCC 511  
Db 181 TTCCCGCGCGGACCCCTTCTCTGAGGGGTCCAGTCCAGGCACCGCGGCTCGGACACCCCC 240  
QY 512 ACCCCCGCGCGGACCTGCTCCCTGGGTGCCCCCTTAACCCGGGCGGTAGCTCGTTAAGATG 571  
Db 241 ACCCCCGCGCGGACCTGCTCCCTGGGTGCCCCCTTAACCCGGGCGGTAGCTCGTTAAGATG 300  
QY 572 GCGAGTGTCCGGTCCGGAACACGCGAAACCCCAATCCCGCTGCCCGACCTCTTGACC 631  
Db 301 GCGAGTGTCCGGTCCGGAACACGCGAAACCCCAATCCCGCTGCCCGACCTCTTGACC 360  
QY 632 CCGCGCCCCACCGGACGACAGACTCGGCTCCCGAGCGCGCGCTCCCGGACACCG 691  
Db 361 CCGCGCCCCACCGGACGACAGACTCGGCTCCCGAGCGCGCGCTCCCGGACACCG 420  
QY 692 GTGCGTGGCAACCGGAGGACCTTTGTAAACGCCACGCTGTTTGTCTTTTGAACAAAG 751  
Db 421 GTGCGTGGCAACCGGAGGACCTTTGTAAACGCCACGCTGTTTGTCTTTTGAACAAAG 480  
QY 752 AATAAATGTGTTAAACTGTCTGAAAGCTTGGCGGCTAAAG 794  
Db 481 AATAAATGTGTTAAACTGTCTGAAAGCTTGGCGGCTAAAG 523



```
RESULT 6
CQ729132/c
LOCUS          4338 bp      DNA      linear      PAT 03-FEB-2004
DEFINITION     Sequence 15066 from Patent WO02068579.
ACCESSION      CQ729132
VERSION        CQ729132.1  GI:42299983
KEYWORDS       Homo sapiens (human)
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1
AUTHORS        Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE          Kits, such as nucleic acid arrays, comprising a majority of
               humanexons or transcripts, for detecting expression and other uses
               thereof
JOURNAL        Patent: WO 02068579-A 15066 06-SEP-2002;
               PE Corporation (NY) (US)
FEATURES       source
               1..4338
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"
ORIGIN
Query Match   12.9%; Score 126; DB 6; Length 4338;
Best Local Similarity 100.0%; Pred. No. 3.1e-16;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCGCGGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGCGCGCGGCGAGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 120
pb 66 ACTGGCGCGCGGCGAGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 7
QY 121 GCCCTG 126
Db 6 GCCCTG 1

RESULT 7
BD244989/c
LOCUS          4338 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION     Mammalian subtilisin/kexin isoenzyme SKI-1: Protein convertase
               having peculiar cleaving characteristic.
ACCESSION      BD244989
VERSION        BD244989.1  GI:33054759
KEYWORDS       Homo sapiens (human)
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 4338)
AUTHORS        Seidah,N., Chretien,M., Marcinkiewicz,M., Laaksonen,R. and
               Davignon,J.
TITLE          Mammalian subtilisin/kexin isoenzyme SKI-1: Protein convertase
               having peculiar cleaving characteristic
JOURNAL        Patent: JP 2002532065-A 3 02-OCT-2002;
               INSTITUT DE RECHERCHES CLINIQUES DE MONTREAL
COMMENT        OS Homo sapiens (human)
               PN JP 2002532065-A/3
               PD 02-OCT-2002
               PF 04-NOV-1999 JP 2000579720
               PR 04-NOV-1998 CA 2249648
               PT NABIL SEIDAH,MICHEL CHRETEN,MIECZYSLAW MARCINKIEWICZ,REIJO
               PI LAAKSONEN,
               PI JEAN DAVIGNON
PC C12N15/09,A61K38/00,A61K45/00,A61P1/16,A61P3/06,A61P9/10, PC
A61P25/28,
PC A61P35/00,A61P43/00,C07K7/06,C07K7/08,C12N1/15,C12N1/19,C12N1/19,

PC 21,C12N5/10,
PC C12N9/50,C12P21/02,C12Q1/02,C12Q1/37,C12Q1/68,G01N33/15,G01N33/ PC
50,
PC G01N33/50,G01N33/566,G01N33/573/(C12N9/50,C12R1:91),C12N15/00, PC
C12N5/00,
PC A61K37/02
CC Mammalian subtilisin/kexin isoenzyme SKI-1:
Protein convertase
CC having
CC peculiar cleaving characteristic
FH key Location/Qualifiers
FT CDS Location/Qualifiers
1..4338
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
FEATURES       source
ORIGIN
Query Match   12.7%; Score 124.4; DB 6; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.7e-16;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCGCGGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGCGCGCGGCGAGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 120
Db 66 ACTGGCGCGCGGCGAGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 7
QY 121 GCCCTG 126
Db 6 GCCCTG 1

RESULT 8
AR261390/c
LOCUS          4338 bp      DNA      linear      PAT 29-JAN-2003
DEFINITION     Sequence 4 from patent US 6322962.
ACCESSION      AR261390
VERSION        AR261390.1  GI:28072400
KEYWORDS       Unknown.
SOURCE         Unknown.
ORGANISM       Unclassified.
REFERENCE      1 (bases 1 to 4338)
AUTHORS        Brown,M.S., Cheng,D., Espenshade,P.J., Goldstein,J.L., Rawson,R.B.
               and Sakai,J.
TITLE          Sterol-regulated Site-1 protease and assays of modulators thereof
JOURNAL        Patent: US 6322962-A 4 27-NOV-2001;
FEATURES       Location/Qualifiers
source         1..4338
               /organism="unknown"
               /mol_type="genomic DNA"
ORIGIN
Query Match   12.7%; Score 124.4; DB 6; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.7e-16;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTCGCGGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db 126 CTCGCGGCGAACAACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 67
QY 61 ACTGGCGCGCGGCGAGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 120
Db 66 ACTGGCGCGCGGCGAGCGCCACAGCTGGGAGCCTCAGCTCCGCGACCCAGCGT 7
QY 121 GCCCTG 126
Db 6 GCCCTG 1
```

```
Db          6  GCCTG 1

RESULT 9
AR527529/c
LOCUS      AR527529          4338 bp      DNA      linear      PAT 08-OCT-2004
DEFINITION Sequence 3 from patent US 6723553.
ACCESSION  AR527529
VERSION     AR527529.1  GI:53914634
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 4338)
AUTHORS   Abrahamse, L., Ekblom, J., Forsgren, M., Horling, J. and Johansson, P.
TITLE     Promoter sequences
JOURNAL   Patent: US 6723553-A 3 20-APR-2004;
FEATURES   Location/Qualifiers
            source
            1..4338
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN
Query Match      12.7%; Score 124.4; DB 6; Length 4338;
Best Local Similarity 99.2%; Pred. No. 6.7e-16;
Matches 125; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  CTCGGCGGCAACACGCCCTGGGCACCTCCATTCCGGGCTGTTTACTCCCAACTCTCGCGAG 60
Db      126  CTCGGCGGCAACACGCCCTGGGCACCTCCATTCCGGGCTGTTTACTCCCAACTCTCGCGAG 67

QY      61  ACTGGCGCGCGCGCGCGAGCGAGCCACACAGCTGGGAGCCTCAGCTCCGCGCGAGCCAGCGCT 120
Db      66  ACTGGCGAGCGGCGCGAGCGAGCGCCACAGCTGGGAGCCTCAGCTCCGCGCGAGCCAGCGCT 7

QY      121  GCCTG 126
Db      6  GCCTG 1

RESULT 11
HUMKIAAL/c
LOCUS      HUMKIAAL          4338 bp      mRNA      linear      PRI 10-JAN-2004
DEFINITION Homo sapiens KIAA0091 mRNA, partial cds.
ACCESSION  D42053
VERSION     D42053.1  GI:577308
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS   Nagase, T., Miyajima, N., Tanaka, A., Suzuki, T., Seki, N., Sato, S.,
          Tabata, S., Ishikawa, K.-I., Kawarabayashi, Y., Kotani, H. and Nomura, N.
          Prediction of the coding sequences of unidentified human genes.
          III. The coding sequences of 40 new genes (KIAA0081-KIAA0120)
          deduced by analysis of cDNA clones from human cell line KG-1
          DNA Res. 2 (1), 37-43 (1995)
          JOURNAL 95308325
          MEDLINE 7788527
          PUBMED
REFERENCE  2 (bases 1 to 4338)
AUTHORS   Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.
TITLE     Direct Submission
JOURNAL   Submitted (08-NOV-1994) Osamu Ohara, Kazusa DNA Research Institute;
          1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
          (E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)
          Location/Qualifiers
            1..4338
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /sex="male"
            /cell_line="KG-1"
            /cell_type="myeloblast"
            1..4338
            /gene="KIAA0091"
            1..496
            /gene="KIAA0091"
            <479..3655

gene
5' UTR
CDS

RESULT 10
AX024613/c
LOCUS      AX024613          4338 bp      DNA      linear      PAT 15-SEP-2000
DEFINITION Sequence 5 from Patent WO0026348.
ACCESSION  AX024613
VERSION     AX024613.1  GI:10184749
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS   Davignon, J., Chretien, M., Laaksonen, R., Seidah, N. and
          Marcinkiewicz, M.
          Mammalian subtilisin/kexin isozyme ski-1: a proprotein
          Patent: WO 0026348-A 5 11-MAY-2000;
          JOURNAL  DAVIGNON JEAN (CA); CHRETIEN MICHEL (CA); LAAKSONEN RELJO (CA);
          SEIDAH NABILA (CA); MARCINKIEWICZ MIECZYSLAW (CA); MONTREAL INST
          RECH CLINIQUES (CA)
          Location/Qualifiers
            1..4338
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            497..3655
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC08995.1"
            /db_xref="GI:10184750"
            /translation="MKLVNIWLLLVLCGKKHGLDRLEKSKFKAPOCGSHLTLK
            VFESTVVEYIVAFNGYFTAKRNSFISSALKSEVDNWRIRIIPNNPSSDPSDFE
            VTQIKKQAGLTLEDHNIKRVTQPKVFRSLKYAESDPTVPCNETRWQWSSR
            PLLRSLSLGSFGFWATGHRSSRRLLRAIPRQVAQTLQADVLWQMGYTGANRVAVFD

FEATURES   Location/Qualifiers
            source
            1..4338
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            497..3655
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC08995.1"
            /db_xref="GI:10184750"
            /translation="MKLVNIWLLLVLCGKKHGLDRLEKSKFKAPOCGSHLTLK
            VFESTVVEYIVAFNGYFTAKRNSFISSALKSEVDNWRIRIIPNNPSSDPSDFE
            VTQIKKQAGLTLEDHNIKRVTQPKVFRSLKYAESDPTVPCNETRWQWSSR
            PLLRSLSLGSFGFWATGHRSSRRLLRAIPRQVAQTLQADVLWQMGYTGANRVAVFD

CDS
```



are a family of serine proteases. They appear to have independently and convergently evolved an Asp/Ser/His catalytic triad, like that found in the trypsin serine proteases (see pfam00089). Structure is an alpha/beta fold containing a 7-stranded parallel beta sheet, order 2314567"  
/db\_xref="CDD:pfam00082"

## ORIGIN

Query Match 9.2%; Score 90.4; DB 9; Length 3026;  
Best Local Similarity 98.9%; Pred. No. 1e-08;  
Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CTCGCGGGAACACACGCTGGGCACTCCATTCCGGGCTGTTTACTCCCAACTCTCCGAG 60  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
92 CTCGCGGGAACACACGCTGGGCACTCCATTCCGGGCTGTTTACTCCCAACTCTCCGAG 33  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 61 ACTGGGCGCGCGGCGAGGAGCCACAGCT 92  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
32 ACTGGGCGCGCGGCGAGGAGCCACAACT 1  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

## RESULT 13

PM11H12G/c  
LOCUS PM11H12G 1393 bp DNA linear STS 29-MAY-2003  
DEFINITION Penicillium marneffei STS, clone pm11h12.g, sequence tagged site.  
ACCESSION AL684264  
VERSION AL684264.1 GI:19338009  
KEYWORDS STS  
SOURCE Penicillium marneffei  
ORGANISM Penicillium marneffei  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
REFERENCE 1  
AUTHORS Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,  
Cai,J.J., Cheung,B.Y., Medigue,C. and Danchin,A.  
TITLE Exploring the Penicillium marneffei genome  
JOURNAL Arch. Microbiol. 179 (5), 339-353 (2003)  
MEDLINE 22595073  
PUBMED 12640520  
REFERENCE 2 (bases 1 to 1393)  
AUTHORS Danchin,A. and Pascal,G.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,  
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong  
FEATURES  
Location/Qualifiers  
1. .1393  
/organism="Penicillium marneffei"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:37727"  
/clone="pm11h12.g"

## ORIGIN

Query Match 6.6%; Score 64.4; DB 11; Length 1393;  
Best Local Similarity 44.4%; Pred. No. 0.0032;  
Matches 288; Conservative 0; Mismatches 349; Indels 11; Gaps 2;  
  
QY 3 CCGCGGCAACAGCTGGGCACTCCATTCCGGGCTGTTTACTCCCAACTCTCCGAGAC 62  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
1212 CCGCGCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1154  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 63 TGGGCGGCGCGGCGAGGAGCCACAGCTGGGAGGCTCAGCTCCGCGGACCCAGGCTGC 122  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
1153 TGGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1094  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 123 CTGTGTCGTCGCGGCTCCCGGGCTTGGGTGCGGCTCTGGACGCGCGTGGGAGCGGGA 182  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
1093 CANNCGGCGCGCGCGCGCGGCTGNCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1034  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 183 CCAGCGCGGAGATGGAGAGAGTGTCTCGGACATTTCG-----GGCGGCGGG 232  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
1033 GGCGGCGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 974  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

233 GCCGTGGCAGGTGGAAACGCGAGGGCGGTGGCCACGAGCTCCAGCGCGCGGAGAACGC 292  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
973 GCG 914  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 293 GCTGGGGGAACCTTTGGTCCGCTCTCGGCGTCCCTTAGAGATCCCGGAAAGGACGG 352  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
913 CCGCTCGGGGCG 854  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 353 GCGCGAAAGCGCGCAGGCTGGGCGAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGC 412  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
853 CCG 794  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 413 CCGCGCGCGCTGCTCTCATGACTCGTGAGAGCTCGCTTCCCGCGCGACCTTCTCTG 472  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
793 CCTCTCNCG 734  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 473 CAGGGGTCCAGTCCAGGCAACCGGCTCGGACACCCACCCCGCGCGGCGGCGGCGGCG 532  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
733 CCG 674  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 533 CTGGGTGCCCTTAACCCGCGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGTCGGAAC 592  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
673 GCG 614  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
QY 593 ACGCGAAACCCAAATCCCGCTGCGGACCTCTGACCCCGCGGCGGCGGCGGCGGCGG 640  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
613 GCGCGCTGCGGCG 566  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
RESULT 14  
AF429315  
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.  
DEFINITION AF429315  
ACCESSION AF429315  
VERSION AF429315.1 GI:17646244  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 125020)  
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,  
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,  
Potter,N.T., Ross,C.A. and Margolis,R.L.  
TITLE A repeat expansion in the gene encoding junctophilin-3 is  
associated with Huntington disease-like 2  
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)  
MEDLINE 21583737  
PUBMED 11694876  
REFERENCE 2 (bases 1 to 125020)  
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.  
TITLE Direct Submission  
JOURNAL Submitted (05-Oct-2001) Psychiatry, Johns Hopkins Medical  
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
FEATURES  
Location/Qualifiers  
1. .125020  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="16"  
/map="16q24.3; between D16S20 and WI-12410"  
/note="isolated from a patient with Huntington's  
Disease-Like 2 (HDL2)"  
complement(35581..35746)  
/rpt\_type=tandem  
/rpt\_unit="ctg"  
gene complement(<36507..>36887)  
/gene="JPH3"  
/note="synonym: JPH3"  
mRNA complement(<36507..>36887)  
/gene="JPH3"  
/product="junctophilin 3"  
CDS complement(<36507..>36887)

```
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSGGRFNFDDGSGYCGWEDGKAHGHGVCCTGPKGQGEYTGWSW
HGFEVLGVYTPSGNTYGTGTAQGRKHGIGLESKGKVVYKGSWTHGPKRGYGVRECA
NGAKYEGTMSNQLQDGYGTETYSYG"

ORIGIN
Query Match          6.6%; Score 64.2; DB 9; Length 125020;
Best Local Similarity 11.6%; Pred. No. 0.0029;
Matches 51; Conservative 260; Mismatches 202; Indels 2; Gaps 1;

QY 44 CTCCTCAACTCTCGAGACTGGCGGCGCGGCGGAGGAGCCACAGTGGAGGCTCAG 103
Db 50678 VDVVBRWBDBTSGWSBHYMRGVHTGVVRWVYGBDTHVYDGGHSGWGMKSCGYSD 50737

QY 104 CTCGCGCAGCCAGCGTCCCTGTCTCCCGCGTCCCGGCGTTCGCTGCGGCTCTG 163
Db 50738 BGKCAKBSKMTCTGSMKTKCTCYGYKWSKADBSYYSBHYBWSYMCCKRBSYASS 50797

QY 164 GACGCGCTGGCGAGCGGAGCCACGCGCGGAGGATGACGAAGTGCTCCGACATTTGG 223
Db 50798 YDVRYKBTGMRYVYRSCSWYVYKYSBMSVSARKSKGHKVBSSMKCSRRKDMSTSSMWR 50857

QY 224 GCGCGCGCGCGCGTGGCAGGCTGGAAGCGGAGGCGGTGGCCAGCGAGCTCCAGCGG 283
Db 50858 SMKDRKCSRSHSYSKSGKWTSDVDRCYACSBMKYKMRKHRYKDSHBSWGSKSHMK 50917

QY 284 CGAGAACGCGCTGGGGAACCTTGCTCGCTGCGCTGCTGCTAGGATCCCGGAAA 343
Db 50918 KGAWRVRSYMRSCSVHDSWVDSRRRMRKGSAMRGMKCYTSSMRSTKSRMSKSWR 50977

QY 344 GGAGCAGCGCGCGAAGCGGCGGAGGCTGGGCGAGGATCTAGAAGACTGCTTGGCGCAG 403
Db 50978 WGSKCYCYGYVWSCTKWRSMGCSYSTGSGSMKMGKYCAGGAGRSRYCKSGSRMSMAGG 51037

QY 404 GCTCCCTGCGCGCGCGGCGTGTCTGATGAGTCTGTGAGAGAGTCTGCTTCCGCGCGGA 463
Db 51038 SMSASMSMGKMTSGRSCARSGRMSYCSRCRCKSKRCSCGCMRCSCWYCWYKCMRYGSM 51097

QY 464 CCTTCTCAGCGGTCCAGCTCAGGCGCGGCGGCTCGGACACCCACCCCGCGCGG 523
Db 51098 RMYKCCMSYKTKYKGCSA--TKSMAMYYSCMRWRKGTSSAWWYSMCCSKCARAASGWY 51155

QY 524 GCACCTGCGCTGGTGCCTTAAACCGCGGCGGTAGCTGCTTAAAG 568
Db 51156 KCMWKGCGCYGGGKGSASTWYMMCCSRACKYTYTTRKSSRGWG 51200

RESULT 15
AC079431 300695 bp DNA linear HTG 01-SEP-2000
LOCUS Mus musculus chromosome 16 clone RP23-1J12, WORKING DRAFT SEQUENCE,
DEFINITION 77 unordered pieces.
AC079431
AC079431.1 GI:9958043
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
```

## COMMENT

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 0  
Center clone name: RPCI-23\_1J12  
-----  
Summary Statistics  
Consensus quality: 164936 bases at least Q40  
Consensus quality: 197223 bases at least Q30  
Consensus quality: 214345 bases at least Q20  
Estimated insert size: 257300; agarose-fp estimation  
Estimated insert size: 293095; sum-of-contigs estimation  
Quality coverage: 3.01 in Q20 bases; agarose-fp estimation  
Quality coverage: 2.65 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 77 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1076: contig of 1076 bp in length  
\* 1077: gap of unknown length  
\* 1177: contig of 1032 bp in length  
\* 2209: gap of unknown length  
\* 2309: contig of 1219 bp in length  
\* 3527: gap of unknown length  
\* 3628: contig of 1137 bp in length  
\* 4765: gap of unknown length  
\* 4865: contig of 1088 bp in length  
\* 5953: gap of unknown length  
\* 6053: contig of 1130 bp in length  
\* 7183: gap of unknown length  
\* 7283: contig of 1138 bp in length  
\* 8421: gap of unknown length  
\* 8521: contig of 1096 bp in length  
\* 9617: gap of unknown length  
\* 9717: contig of 1413 bp in length  
\* 11130: gap of unknown length  
\* 11229: gap of unknown length  
\* 11230: contig of 1179 bp in length  
\* 12409: gap of unknown length  
\* 12509: contig of 1259 bp in length  
\* 13768: gap of unknown length  
\* 13868: contig of 1177 bp in length  
\* 15045: gap of unknown length  
\* 15145: contig of 1181 bp in length  
\* 16326: gap of unknown length  
\* 16426: contig of 1153 bp in length  
\* 17579: gap of unknown length  
\* 17679: contig of 1167 bp in length  
\* 18846: gap of unknown length  
\* 18946: contig of 1115 bp in length  
\* 20061: gap of unknown length  
\* 20161: contig of 1098 bp in length  
\* 21259: gap of unknown length  
\* 21359: contig of 1032 bp in length  
\* 22391: gap of unknown length  
\* 22491: contig of 1109 bp in length  
\* 23600: gap of unknown length  
\* 23700: contig of 1208 bp in length  
\* 24908: gap of unknown length  
\* 26121: contig of 1114 bp in length  
\* 26222: gap of unknown length  
\* 27274: contig of 1052 bp in length  
\* 27374: gap of unknown length  
\* 28472: contig of 1099 bp in length  
\* 28473: gap of unknown length  
\* 28573: contig of 1111 bp in length



Qy 643 GCGACGACAG 652  
Db 13437 CCGNNGCCG 13446

Search completed: October 9, 2005, 02:21:59  
Job time : 4302 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 22:51:27 ; Search time 7895 Seconds  
(without alignments)

11903.823 Million cell updates/sec

Title: US-10-829-118-2

Perfect score: 2469

Sequence: 1 ttgagctctctggaggctc.....gaatacatggcctacagct 2469

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:\*

1: gb\_est1:.\*  
2: gb\_est2:.\*  
3: gb\_hic:.\*  
4: gb\_est3:.\*  
5: gb\_est4:.\*  
6: gb\_est5:.\*  
7: gb\_est6:.\*  
8: gb\_gss1:.\*  
9: gb\_gss2:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	471.6	19.1	500	6	CB049709 NISC_gj12
C 2	468	19.0	468	1	AI792242 ov05e09.y
C 3	451.8	18.3	455	1	AI760651 w166g09.x
C 4	448	18.1	450	2	AW051661 wx27b06.x
C 5	440.6	17.8	1240	5	B0437878 AGENCOURT
C 6	435	17.6	435	1	AI685715 tu37c07.x
C 7	435	17.6	435	1	AI733521 ov05e09.x
C 8	435	17.6	435	2	BF446188 7p33f05.x
C 9	432	17.5	432	1	AI554846 te45b04.x
C 10	430.2	17.4	436	2	AW005919 w290e09.x
C 11	411.8	16.7	415	1	AI362974 qy87h07.x
C 12	404	16.4	404	2	BE048839 hr54e04.x
C 13	362.8	14.7	389	1	AI498080 tm66a04.x
C 14	332	13.4	3491	3	BC039100 Homo sapi
C 15	305.2	12.4	897	6	CD557847 AGENCOURT
C 16	301.6	12.3	457	2	BF570483 602186293
C 17	299.6	12.1	15970	8	A0839852 260113-C5
C 18	288.4	12.1	881	5	BUS21286 AGENCOURT
C 19	280.4	11.8	736	6	CA427039 UI-H-DF0-
C 20	282.2	11.7	2786	3	BC035101 Homo sapi
C 21	287.6	11.6	918	5	BQ706343 AGENCOURT
C 22	282.4	11.4	672	6	CA431692 UI-H-DF0-
C 23	280.8	11.4	3180	3	HSM800699 Homo sapi
C 24	280.6	11.4	2097	3	CR614786 full-leng

25	279.4	11.3	630	5	BX508391	BX508391 DKFP686F
26	277.4	11.2	629	5	BX509360	BX509360 DKFP686J
27	277.2	11.2	778	1	AV764490	AV764490 AV764490
C 28	276	11.2	3474	3	HSM800886	AL110229 Homo sapi
C 29	274.8	11.1	834	8	BZ611716	BZ611716 WHAA044TR
C 30	273.8	11.1	1027	1	AV762129	AV762129 AV762129
C 31	273.8	11.1	1027	5	BQ070852	BQ070852 AGENCOURT
C 32	273.8	11.1	1036	5	BQ070834	BQ070834 AGENCOURT
C 33	273.6	11.1	2558	3	BC038553	BC038553 Homo sapi
C 34	273.4	11.1	3016	3	BC068461	BC068461 Homo sapi
C 35	273.2	11.1	4828	3	HSM802759	AL390128 Homo sapi
36	272.6	11.0	5479	3	HSM807678	BX647532 Homo sapi
37	272	11.0	610	7	CR540737	CR540737 DKFP459G
38	271.8	11.0	923	6	CD518127	CD518127 AGENCOURT
39	271	11.0	910	5	BQ722917	BQ722917 AGENCOURT
40	270.8	11.0	3840	3	BC032939	BC032939 Homo sapi
41	270	10.9	3566	3	HSM805040	AL833727 Homo sapi
42	269.8	10.9	694	7	CR764988	CR764988 DKFP470H
43	269.2	10.9	680	8	AQ356404	AQ356404 CTBT1-E1-
44	268.8	10.9	660	5	BX484570	BX484570 DKFP686I
45	268.8	10.9	769	5	BQ710315	BQ710315 AGENCOURT

#### ALIGNMENTS

RESULT 1  
CB049709/c  
LOCUS  
DEFINITION  
3', mRNA sequence.  
CB049709  
CB049709.1 GI:27787996  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 500)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
TITLE  
Unpublished (1997)  
JOURNAL  
Contact: Robert Strausberg, Ph.D.  
COMMENT  
Email: cgaabs-r@mail.nih.gov  
CDNA Library Preparation:  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
info@image.llnl.gov  
Plate: LLAM8008 row: L column: 2  
Seq primer: -21M13 forward primer (ABI).  
Location/Qualifiers  
1. 500  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3271441"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="NCI-CGAP Pr28"  
/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI-CGAP Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonoids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 19.1%; Score 471.6; DB 6; Length 500;  
Best Local Similarity 99.2%; Pred. No. 6.5e-66;  
Matches 474; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 652 GAACCTTGGTCCGCTCTGCGCTGCTCTAGGATCCCGAAAGAGGACACGGCGCGAA 711  
DB 495 GAGGCTTGGTCCGCTCTGCGCTGCTCTAGGATCCCGAAAGAGGACACGGCGCGAA 436  
QY 712 AGCGCCAGGCTGGCCAGAGATCTAGAAAGACTGCTTGGGCGAGGCTCCCTGCCCGCGG 771  
DB 435 AGCGCCAGGCTGGCCAGAGATCTAGAAAGACTGCTTGGGCGAGGCTCCCTGCCCGCGG 376  
QY 772 GGCTGTGTGTCATGAGCTGTGAGAGCTGCTTCCGCGCGACCTTCTCTCAGGGGT 831  
DB 375 GGCTGTGTGTCATGAGCTGTGAGAGCTGCTTCCGCGCGACCTTCTCTCAGGGGT 316  
QY 832 CCAGCTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTGCCTCGGGTG 891  
DB 315 CCAGCTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTGCCTCGGGTG 256  
QY 892 CCCTTAAACCGCGGCTAGCTGTTAAGATGGGAGTGTCCGTCGCGAAACACGCGAA 951  
DB 255 CCCTTAAACCGCGGCTAGCTGTTAAGATGGGAGTGTCCGTCGCGAAACACGCGAA 196  
QY 952 ACCCGAAATCCCGCTGCCGACTCTGACCCCGCGCGGCGACACTGGGCG 1011  
DB 195 ACCCGAAATCCCGCTGCCGACTCTGACCCCGCGCGGCGACACTGGGCG 136  
QY 1012 CTCCGACGCGGACGCGCTGCCGCGACACCGGTGGTGGAAACGAGGACCTTTGTAA 1071  
DB 135 CTCCGACGCGGACGCGCTGCCGCGACACCGGTGGTGGAAACGAGGACCTTTGTAA 76  
QY 1072 CGCACGTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAA 1129  
DB 75 CGCACGTGTTGCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTCTGAAA 18

## RESULT 2

AI792242  
LOCUS ov05e09.y5 NCI\_CGAP\_kid3 Homo sapiens cDNA clone IMAGE:1636456 5',  
DEFINITION mRNA sequence.  
ACCESSION AI792242.1 GI:53399949  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 468)  
TITLE NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute / National Institute of Dental Research,  
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Other\_ESTS: ov05e09.x5  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-i@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
This read is a RESEQUENCE of a previously sequenced human clone  
Original clone citation: see original entry for original citation  
Information  
This 5' resequenced clone has no previous 5' data to verify this  
new read against

Putative full length read  
The vector to vector length is 469  
Seq primer: -40RP from Gibco.

## FEATURES

source  
1..468  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1636456"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_kid3"  
/note="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer,  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified p7T3 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo."

## ORIGIN

Query Match 19.0%; Score 468; DB 1; Length 468;  
Best Local Similarity 100.0%; Pred. No. 2.5e-65;  
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 CTTTGGTCCGCTCGCGCTGCTCTAGGATCCCGAAAGAGGACACGGCGCGAAAGCG 715  
DB 1 CTTTGGTCCGCTCGCGCTGCTCTAGGATCCCGAAAGAGGACACGGCGCGAAAGCG 60  
QY 716 GCCAGGCTGGGCGCAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCCTGCCCGCGGGGC 775  
DB 61 GCCAGGCTGGGCGCAGGATCTAGAAAGACTGCTTGGCGCAGGCTCCCTGCCCGCGGGGC 120  
QY 776 TGTCTGTATGTAATGCTGTGAGAGCTGCTTCCGCGCGGACCTTCTCTGAGGGGTCCAC 835  
DB 121 TGTCTGTATGTAATGCTGTGAGAGCTGCTTCCGCGCGGACCTTCTCTGAGGGGTCCAC 180  
QY 836 GTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTCTCTGAGGGGTCCAC 895  
DB 181 GTCCAGGACCGCGGCTCGGACACCCACCCCGCGCGGACCTTCTCTGAGGGGTCCAC 240  
QY 896 TTAACCCCGCGGCTAGCTGTTAAGATGCGAAGTGTCCGTCGCGGAAACACGCGAAACCC 955  
DB 241 TTAACCCCGCGGCTAGCTGTTAAGATGCGAAGTGTCCGTCGCGGAAACACGCGAAACCC 300  
QY 956 CAAATCCCGCTTCCCGACCTCTGACCCCGCGCGGCGGACGACAGACTGGGGCTCC 1015  
DB 301 CAAATCCCGCTTCCCGACCTCTGACCCCGCGCGGCGGACGACAGACTGGGGCTCC 360  
QY 1016 CGACGCGGACGCGCTGCCGCGACACCGGTGGTGGGAAACGAGGACCTTTGTAAAGCC 1075  
DB 361 CGACGCGGACGCGCTGCCGCGACACCGGTGGTGGGAAACGAGGACCTTTGTAAAGCC 420  
QY 1076 ACCTGTTTCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTC 1123  
DB 421 ACCTGTTTCTCTTTTGAAGAAACAAAGAAATAAATGTGTTAACTGTC 468

## RESULT 3

AI760651/c  
LOCUS AI760651.x1 NCI\_CGAP\_Kid12 Homo sapiens cDNA clone IMAGE:2398336 3',  
DEFINITION mRNA sequence.  
ACCESSION AI760651  
VERSION AI760651.1 GI:5176318  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 455)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
**JOURNAL** Tumor Gene Index  
**COMMENT** Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image.html  
Insert Length: 567 Std Error: 0.00  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. .455  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2398336"  
/tissue\_type="2 pooled tumors (clear cell type)"  
/lab\_host="DH10B"  
/notes="Organ: Kidney; Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

**FEATURES**  
source  
1. .455  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2398336"  
/tissue\_type="2 pooled tumors (clear cell type)"  
/lab\_host="DH10B"  
/notes="Organ: Kidney; Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

**ORIGIN**  
Query Match 18.3%; Score 451.8; DB 1; Length 455;  
Best Local Similarity 99.6%; Pred. No. 1e-62; Indels 0; Gaps 0;  
Matches 453; Conservative 0; Mismatches 2;  
675 TCCTCTTAGGATCCCCGAAAGGAGCAGCGGCGCGAAAGCGGCGAGCTGGCGCAGATC 734  
675 TCCTCTTAGGATCCCCGAAAGGAGCAGCGGCGCGAAAGCGGCGAGCTGGCGCAGATC 396  
735 TAGAAGACTGCTGGCGCAGGCTCCCTGCCCCCGGGGCTGCTGTCATGACTCGTGG 794  
395 TAGAAGACTGCTGGCGCAGGCTCCCTGCCCCCGGGGCTGCTGTCATGACTCGTGG 336  
795 AGAGCTCGCTTCCCGCGGACCCCTCTCGAGGGTCCACGTCAGGCACCGCGGCTC 854  
335 AGAGCTCGCTTCCCGCGGACCCCTCTCGAGGGTCCACGTCAGGCACCGCGGCTC 276  
855 GGACACCCCAACCCCGCGCGGACCTGCGCTGGGTGCCCTTAAACCCGGGCGGTAGCTC 914  
275 GGACACCCCAACCCCGCGCGGACCTGCGCTGGGTGCCCTTAAACCCGGGCGGTAGCTC 216  
915 GTTAAGATGGCGAAGTGTCCGGTCCGGAACACCGCGAAACCCCAATCCCGCTGCCCGAC 974  
215 GTTAAGATGGCGAAGTGTCCGGTCCGGAACACCGCGAAACCCCAATCCCGCTGCCCGAC 156  
975 CTCTGACCCCGCGCCACGGGACGACACTGGGCTCCCGACGGCGCGGCTGCC 1034  
155 CTCTGACCCCGCGCCACGGGACGACACTGGGCTCCCGACGGCGCGGCTGCC 96  
1035 GGACACCGGTGCGTGGAAACGGAGGACCTTTTGTAAACCGCACCTGTTGCTCTTTTGA 1094  
95 GGACACCGGTGCGTGGAAACGGAGGACCTTTTGTAAACCGCACCTGTTGCTCTTTTGA 36  
1095 AAAAACGAATAAATGTGTTAAATCTGTCTGAAAA 1129  
35 AAAAACGAATAAATGTGTTAAATCTATCTGAAAA 1

**RESULT 4**  
**LOCUS** AW051661/c  
**DEFINITION** wx27b06.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2544851 3', mRNA sequence.  
**ACCESSION** AW051661  
**VERSION** AW051661.1 GI:5913892  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 450)  
**AUTHORS** NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image.html  
Insert Length: 540 Std Error: 0.00  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. .450  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2544851"  
/lab\_host="DH10B"  
/notes="Organ: Kidney; Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 132376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

**ORIGIN**  
Query Match 18.1%; Score 448; DB 2; Length 450;  
Best Local Similarity 100.0%; Pred. No. 4.1e-62; Indels 0; Gaps 0;  
Matches 448; Conservative 0; Mismatches 0;  
676 CGCTCTAGGATCCCCGAAAGGAGCAGCGGCGCGAAAGCGGCGAGCTGGCGCAGATCT 735  
450 CGCTCTAGGATCCCCGAAAGGAGCAGCGGCGCGAAAGCGGCGAGCTGGCGCAGATCT 391  
736 AGAAGACTGCTGGCGCAGGCTCCCTGCCCCCGGGGCTGCTGTCATGACTCGTGGGA 795  
390 AGAAGACTGCTGGCGCAGGCTCCCTGCCCCCGGGGCTGCTGTCATGACTCGTGGGA 331  
796 GAGCTGCTTCCCGCGGACCCCTTCTGAGGGGTCCAGTCCAGGCACCGCGGCTCG 855  
330 GAGCTGCTTCCCGCGGACCCCTTCTGAGGGGTCCAGTCCAGGCACCGCGGCTCG 271  
856 GACACCCCAACCCCGCGCGGACCTGCTCCCTGGGTGCCCTTAAACCCGGGCGGTAGCTCG 915  
270 GACACCCCAACCCCGCGCGGACCTGCTCCCTGGGTGCCCTTAAACCCGGGCGGTAGCTCG 211  
916 TTAAGATGGCGAAGTGTCCGTCGGGAACACCGGAAACCCCAATCCCGCTGCCCGACC 975  
210 TTAAGATGGCGAAGTGTCCGTCGGGAACACCGGAAACCCCAATCCCGCTGCCCGACC 151

```

QY 976 TCCTGACCCCGCCCGCCAGACGAGACTGGGCTCCGCGAGCGCGCTGCCG 1035
DB 150 TCTTGACCCCGCCCGCCAGACGAGACTGGGCTCCGCGAGCGCGCTGCCG 91
QY 1036 GGACACCGGTGCTGCGAAACGAGACCTTTGTAACGCCACGTGTTGCTCTTTTGAA 1095
DB 90 GGACACCGGTGCTGCGAAACGAGGACCTTTGTAACGCCACGTGTTGCTCTTTTGAA 31
QY 1096 AAAACAAGATTAATGTTAACTGTC 1123
DB 30 AAAACAAGATTAATGTTAACTGTC 3

RESULT 5
BQ437878 1240 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7890066 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6176570
DEFINITION 5', mRNA sequence.
ACCESSION BQ437878
VERSION BQ437878.1 GI:21176954
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1240)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13553 row: C column: 03
High quality sequence stop: 421.
Location/Qualifiers
  1..1240
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:6176570"
  /tissue_type="leiomyosarcoma"
  /lab_host="DH10B (phage-resistant)"
  /clone_lib="NIH MGC 71"
  /notes="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

FEATURES
  source
    Query Match 17.8%; Score 440.6; DB 5; Length 1240;
    Best Local Similarity 94.4%; Pred.No. 4.9e-61;
    Matches 490; Conservative 0; Mismatches 24; Indels 5; Gaps 3;

QY 626 TGCCAGCGCGGAGAACGCGCTCGGGGAAACCTTGGTCCGCTCGCGCGTCTAGGA 685
DB 1 TGCCAGCGCGGAGAACGCGCTCGGGGAAACCTTGGTCCGCTCGCGCGTCTAGGA 60
QY 686 TCCCCGAAAGAGAGCAGCGCGCGGAAAGCGGCCAGGATCTAGAAAGACTG 745
DB 61 TCCCCGAAAGAGAGCAGCGCGCGGAAAGCGGCCAGGATCTAGAAAGACTG 120
QY 746 CTTGGCGCAGGCTCCCTGCCCCCGCGGCTGTGTATGACTCGTGGAGAGCTGCTT 805
DB 121 CTTGGCGCAGGCTCCCTGCCCCCGCGGCTGTGTATGACTCGTGGAGAGCTGCTT 180
QY 806 CCGCGCGGAGACCTTCTCGAGGGGTCACGTCAGGACCGCGCGCTCGACACCCAC 865
DB 181 CCGCGCGGAGACCTTCTCGAGGGGTCACGTCAGGACCGCGCGCTCGACACCCAC 240

```

```

QY 866 CCCCGCGCGGACACCTGCGCTGCCCTTAAACCCGCGCGGTAGCTCGTTAAGATGCC 925
DB 241 CCCCGCGCGGACACCTGCGCTGCCCTTAAACCCGCGCGGTAGCTCGTTAAGATGCC 300
QY 926 GAAGTGTCCGGTCCCGAACACGCGAAACCCAAAATCCCGCTGCCGACCTCTGTGACCCC 985
DB 301 GAAGTGTCCGGTCCCGAACACGCGAAACCCAAAATCCCGCTGCCGACCTCTGTGACCCC 360
QY 986 CGGCGCCACGCGACGACAGACTGGGCTCCGACGCGGAGCGGCTGCCGCGGACACC-GG 1044
DB 361 CGGCGCCACGCGACGACAGACTGGGCTCCGACGCGGAGCGGCTGCCGCGGACACC-GG 420
QY 1045 TCGGTGCGGAAACGG--AGGACCTTTGTAAACGCGACGCTG--TTTGTCTCTTTTGAAGAAC 1100
DB 421 TCGGTGCGGAAACGGAGGACCTTTGTAAACGCGACGCTGCTGCTCTTTTGAAGAAC 480
QY 1101 AAGAATAAATGTTAAACTGTCTGAAAGCTTGGCGCC 1139
DB 481 CAGAATAATGCGCTCACTGGCTGAAAAGGTCGTGCC 519

RESULT 6
AI685715 435 bp mRNA linear EST 07-MAR-2000
LOCUS tu37c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253228 3',
DEFINITION mRNA sequence.
ACCESSION AI685715
VERSION AI685715.1 GI:4897009
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 562 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
  1..435
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:2253228"
  /sex="male"
  /dev_stage="adult"
  /lab_host="DH10B"
  /clone_lib="NCI CGAP Pr28"
  /notes="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and as
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonesIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

FEATURES
  source
    Query Match 17.6%; Score 435; DB 1; Length 435;
    ORIGIN

```











QY 715 GGCCAGGCTGGCCAGGATCTAGAAAGACTGCTGCGGCGAGGCTCCCTGCCCGCGGGC 774  
DB 415 GGCCAGGCTGGCCAGGATCTAGAAAGACTGCTGCGGCGAGGCTCCCTGCCCGCGGGC 356  
QY 775 CTGCTGTCTGAGTCTGCTGAGAGCTCGCTTTCCCGCGGACCCCTTCTTCCAGGGTCCA 834  
DB 355 CTGCTGTCTGAGTCTGCTGAGAGCTCGCTTTCCCGCGGACCCCTTCTTCCAGGGTCCA 296  
QY 835 CGTCCAGGACCCGGCGGCTCGGACACCCCAACCCCGCGGGCACTTGCCTGGGTGCC 894  
DB 295 CGTCCAGGACCCGGCGGCTCGGACACCCCAACCCCGCGGGCACTTGCCTGGGTGCC 236  
QY 895 CTTAAACCCGGCGGCTGAGTCTGAGTGGCGAGGCTCGGTCGGGACACCGGAAAC 954  
DB 235 CTTAAACCCGGCGGCTGAGTCTGAGTGGCGAGGCTCGGTCGGGACACCGGAAAC 176  
QY 955 CCAAAATCCCGGCTCGGACCTCTGACCCCGCGGCGGACGACAGACTGGGCTC 1014  
DB 175 CCAAAATCCCGGCTCGGACCTCTGACCCCGCGGCGGACGACAGACTGGGCTC 116  
QY 1015 CGGACCGGACCGCTGCGGAGCACCGGTGCGTGGGAAACCGAGGACCTTTGTAAAGC 1074  
DB 115 CGGACCGGACCGCTGCGGAGCACCGGTGCGTGGGAAACCGAGGACCTTTGTAAAGC 56  
QY 1075 CAGCTGTTGCTCTTTTGGAAAAACAGAAATGTTGTTAACTGTCTGAAAA 1129  
DB 55 CAGCTGTTGCTCTTTTGGAAAAACAGAAATGTTGTTAACTGTCTGAAAA 1

## RESULT 12

BE048839/c  
LOCUS hr54e04.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3132318 3',  
DEFINITION mRNA sequence.

ACCESSION BE048839

VERSION BE048839.1 GI:8365894

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 404)

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

EMAIL: cgapbs-r@mail.nih.gov

TISSUE Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Sequencing by: Washington University Genome Sequencing Center

Cloning Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. .404

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3132318"

/lab\_host="DH10B"

/clone\_lib="NCI CGAP Kid11"

/notes="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Plasmid DNA from the normalized library NCI\_CGAP\_Kid11 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

## FEATURES

source

## ORIGIN

Query Match 16.4%; Score 404; DB 2; Length 404;  
Best Local Similarity 100.0%; Pred. No. 4.9e-55;  
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 722 CTGGGCGAGGATCTAGAAAGACTGCTGCGGCGAGGCTCCCTGCCCGCGGGCTGCTGT 781

DB 404 CTGGGCGAGGATCTAGAAAGACTGCTGCGGCGAGGCTCCCTGCCCGCGGGCTGCTGT 345

QY 782 CATGAGCTCGTGGAGAGCTCGCTTCCCGCGGAGCCCTTCTTCCGAGGGTCCACGTCAG 841

DB 344 CATGAGCTCGTGGAGAGCTCGCTTCCCGCGGAGCCCTTCTTCCGAGGGTCCACGTCAG 285

QY 842 GCACCGCGGCTCGGACACCCACCCCGCGGCGGACCTTGGCTGGCTGCTTAAACC 901

DB 284 GCACCGCGGCTCGGACACCCACCCCGCGGCGGACCTTGGCTGGCTGCTTAAACC 225

QY 902 CGGCGGCTAGCTGCTTAAAGTGGCGAAGTGTCCGCTCCGGAACACCGCAAAACCCCAATC 961

DB 224 CGGCGGCTAGCTGCTTAAAGTGGCGAAGTGTCCGCTCCGGAACACCGCAAAACCCCAATC 165

QY 962 CGGCTGCGCGGACCTCTGACCCCGCGGCGGACGACGACTGGGCTCCCGACGC 1021

DB 164 CGGCTGCGCGGACCTCTGACCCCGCGGCGGACGACGACTGGGCTCCCGACGC 105

QY 1022 GCAGCGGCTGCGGCGGACACCGGTGCGTGGCGAAACCGAGGACCTTTGTAAACGCGCTGT 1081

DB 104 GCAGCGGCTGCGGCGGACACCGGTGCGTGGCGAAACCGAGGACCTTTGTAAACGCGCTGT 45

QY 1082 TTGCTCTTTTGAAGAAACCAAGAAATGTTGTTAACTGTCTG 1125

DB 44 TTGCTCTTTTGAAGAAACCAAGAAATGTTGTTAACTGTCTG 1

## RESULT 13

AI498080/c

LOCUS

DEFINITION

AI498080

VERSION

AI498080.1 GI:4390062

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 389)

TITLE NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

COMMENT National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BTGP), Tumor Gene Index

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Sequencing by: Greg Lennon, Ph.D.

Cloning Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 516 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 378.

Location/Qualifiers

1. .389

/organism="Homo sapiens"

source



Db 1430 GGAGGTTGAGGAGCGGAGATTGGCCACTGCGCTCTGCTGGGTGACAGAGTGAGAC 1489

Qy 2164 CTGTCTCAAAAACAAAACAAAAGCATCTATAAA 2199

Db 1490 CCGTCTCAAAATATAATAATAATAATAATAAA 1525

RESULT 15

CD557847 897 bp mRNA linear EST 11-JUN-2003

LOCUS AGENCOURT\_14423258 NIH\_MGC\_180 Homo sapiens cDNA clone

DEFINITION IMAGE:30390116 5', mRNA sequence.

ACCESSION CD557847

VERSION CD557847.1 GI:31583915

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 897)

REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL

COMMENT Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
cDNA Library Preparation: Dr. Michael Brownstein  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
plate: NDAM464 row: k column: 21  
High quality sequence stop: 636.  
Location/Qualifiers  
1..897  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30390116"  
/lab\_host="DH10B-Ton A (T1 and T5 phage resistances)"  
/notes="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH\_MGC Library."

FEATURES  
source

ORIGIN

Query Match 12.4%; Score 305.2; DB 6; Length 897;  
Best Local Similarity 68.9%; Pred. No. 3e-39;  
Matches 495; Conservative 0; Mismatches 208; Indels 15; Gaps 5;

Qy 1472 TTGTTTAACTCTTTGTACCTCAGCTTCTTCATTACAAAATAGGGGTACTAGCCAGGCG 1531

Db 34 TTGTGTTTCTCACTACATTGTTACATGTGGGAATTACAGATAACGGAAGCCGGCTGGG 93

Qy 1532 GGGTGGCTCGCGCTGTATCCAGACATTTGGGAGGCCGAGCAGCCGGATCACTTGAG 1591

Db 94 TGGTGGCTCACGGCTGTAAACCCCAACACTTTGGGAGGCCAAGCGCGGATCACCTGAG 153

Qy 1592 GTCAGAAAGTTTCAGACAGCTGTCAACATGGGTGAACCGCGCTCTACCAAAAATAT 1651

Db 154 GTCAGAGTTTCAGAGATTAGTCTGGCCAAATCAT-GGTGAACCCCATCTCTACTAAAAATAC 212

Qy 1652 AAAAACTTAGCTGAGTGTGGTAGCGCATGCTGTAATCCAGCAACTCAGGAGGCTGAGG 1711

Db 213 GAAA--TTAGCCAGGTGTGGTGGCACACATCTGTGTGCTCCAGCTACTCTGAGGCTGAGA 270

Qy 1712 CA-GAAGATCGCTTGAACTCTGGAGCGGAGGTTGCAGTGAGCTGAGATCGTACCACTGC 1770

Db 271 CAGGAGATCGCTTGAACTCCAGGAAGTGGAGGTTGCAGTGAGCTGAGATCACCACTGC 330

Qy 1771 ACTCCAGCTTTGGCGCAGAGCGAGACTCTGCTCTAAATAATAATAATAATAATAA 1830

Db 331 ACTCCAGCTTTGGCGCAGAGCGAGAGTGTCTCCATCTCAAAAAA-----AAAA 381

Qy 1831 AAAATAGGGTACTAATATCTACCTTAAAGGATGAGGTTAAATTAAGTACACACATAAG 1890

Db 382 GATAGAAGCAATAAGCATGGTGCAATCAAAATTCTGGCAAGCATTAATAATCAGGATGCA 441

Qy 1891 CCTAGCGCAGTGGCTTATGCTGTAATCTCAACACTTTGGGAGTCTGTGGCGGAGGAT 1950

Db 442 GCTGGCAGCGTGGCTCACGCTGTAAATCCAGCACTTTGGGAGGCAAGGTGGGAGAT 501

Qy 1951 CACTTGAGCCAGGAGTTTGAGACTAGTCTGGGCAACAGAGACATGTCTCTATAGT--TG 2008

Db 502 CACTTGAGGTCAAGAAATTTGAGAGGAGCGCTGGCCAGCATGGCAAAACCCCATCTGTACTT 561

Qy 2009 TGTGTTGGTTTGTGTTTACAGGTGTGGTGTGTCACCTGCAGTCCCGAGCTACTAGGA 2068

Db 562 AAAATACAAAAAATTAGCTGGGCGTGTGTGCACACCTGTAATCCAGCTACTTTGGGA 621

Qy 2069 GGCTGAGGTGGGAGGACTGCTGAGCCAGGAGTCCAGGCTGAGTGCAGTGCATGTTGT 2128

Db 622 GGCTGAGGTGGGAGAAATTTGTAACCTGGAGGTGGAGTTGCAGTGAGCTGAGATCCT 681

Qy 2129 GCCACTGCACTCCAGCTGGGCAACACAGCAAGACCTTGTCTCAAAAAACAACAAAA 2186

Db 682 GCCACTGCACTCCAGGCTGGGCAACAGAGTGAGACCATGTCTCAAAAAATAAAAAATA 739

Search completed: August 12, 2005, 04:53:20  
Job time : 7901 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 12, 2005, 02:41:53 ; Search time 1559 Seconds  
(without alignments)  
10275.313 Million cell updates/sec

Title: US-10-829-118-2

Perfect score: 2469

Sequence: 1 ttgagctctctggaggctc.....gaatacatggccctacagct 2469

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7305758 seqs, 3244068913 residues

Total number of hits satisfying chosen parameters: 14611516

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 22: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 23: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 24: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 25: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 26: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2469	100.0	2469	9	US-09-891-711-2
2	2469	100.0	2469	9	US-10-829-118-2
3	980	39.7	980	9	US-09-891-711-1
4	980	39.7	980	19	US-10-829-118-1
5	332	13.4	430442	20	US-10-417-375-128
6	330.8	13.4	100445	19	US-10-322-281-170
7	327.4	13.3	147620	20	US-10-723-860-2768

8	324.4	13.1	439892	13	US-10-087-192-454
9	323.6	13.1	344805	20	US-10-779-271-1
10	323.6	13.1	354592	22	US-10-737-082-70
11	323.6	13.1	354592	22	US-10-765-790-70
12	322.6	13.1	1980090	20	US-10-719-993-6815
13	322.6	13.1	1980090	21	US-10-741-600-17676
14	322.2	13.0	105413	17	US-10-427-923-3
15	322.2	13.0	91760	13	US-10-087-192-844
16	321.4	13.0	47219	22	US-10-699-941-10
17	321.4	13.0	59884	22	US-10-699-941-11
18	321	13.0	12319	20	US-10-719-993-6911
19	321	13.0	12319	21	US-10-741-600-17830
20	320.4	13.0	28693	21	US-10-741-600-17761
21	319.2	12.9	94720	18	US-10-052-482-160
22	318	12.9	23208	21	US-10-741-600-17814
23	315.6	12.8	170834	9	US-09-835-232-7
24	315.6	12.8	170834	16	US-10-308-485-7
25	314.6	12.7	49195	21	US-10-741-600-17983
26	314.4	12.7	72691	21	US-10-741-600-17712
27	314.4	12.7	109906	18	US-10-235-192A-31
28	313.8	12.7	133893	13	US-10-161-510-1
29	313.4	12.7	53623	20	US-10-417-375-44
30	313.2	12.7	1242	13	US-10-027-632-123227
31	313.2	12.7	1242	17	US-10-027-632-123227
32	313	12.7	304905	18	US-10-271-416-1
33	312	12.6	1250	13	US-10-027-632-249398
34	312	12.6	1250	17	US-10-027-632-249398
35	312	12.6	10770	17	US-10-074-024-515
36	312	12.6	12613	17	US-10-074-024-514
37	312	12.6	87687	13	US-10-087-192-1978
38	311.8	12.6	82121	17	US-10-085-117-136
39	311.6	12.6	32249	10	US-09-764-891-8024
40	310.8	12.6	21721	18	US-10-650-507-41
41	310.6	12.6	330926	20	US-10-719-993-7048
42	310.2	12.6	22976	18	US-10-650-507-19
43	310.2	12.6	23187	16	US-10-214-684A-1
44	310.2	12.6	41936	9	US-09-967-768A-116
45	310.2	12.6	41936	21	US-10-843-641A-6261

ALIGNMENTS

RESULT 1

US-09-891-711-2  
; Sequence 2, Application US/09891711  
; Patent No. US20020082404A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Promoter Sequences  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/09/891,711  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2469  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-891-711-2

Query Match 100.0%; Score 2469; DB 9; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TTGAGTCTGTCTGGAGCTCCGGCCAGACAGAGCGGGCGTATTGTTTCACTCGGTGAATGCT	60
Db	1	TTGAGTCTGTCTGGAGCTCCGGCCAGACAGAGCGGGCGTATTGTTTCACTCGGTGAATGCT	60
Qy	61	CATTTCACGTAAAGAAACCCAGCAACGGAACGAGCTCCGGAGCCGCGAGACCCCGCA	120
Db	61	CATTTCACGTAAAGAAACCCAGCAACGGAACGAGCTCCGGAGCCGCGAGACCCCGCA	120

121 GGGCCGGGTACAGGCACGCTGTGTCACAAACAGCGCGAGGCCCGCGCCACCTCC 180  
121 GGGCCGGGTACAGGCACGCTGTGTCACAAACAGCGCGAGGCCCGCGCCACCTCC 180  
181 CCGACCCGGCCCGGCCCGCCCGAGCCCTCGCTCGGGGCTTCGAGCAACCGCACACCT 240  
181 CCGACCCGGCCCGGCCCGCCCGAGCCCTCGCTCGGGGCTTCGAGCAACCGCACACCT 240  
241 GAGCGAGCGGCCCGCCACCGCTAGGCGGAGCGGCTCGGGAGGCCCGCGCGCGGCTG 300  
241 GAGCGAGCGGCCCGCCACCGCTAGGCGGAGCGGCTCGGGAGGCCCGCGCGCGGCTG 300  
301 ACCTACTCGCGCCCGCGGAGCTCAGGGCCGGCGCGGCGCGGGATGACGGCGCTCCGCG 360  
301 ACCTACTCGCGCCCGCGGAGCTCAGGGCCGGCGCGGCGCGGGATGACGGCGCTCCGCG 360  
361 CGAACACGCTCGGCGACTCCATTCGGGGCTGTCTTACTCCCACTCTCGCGAGACTGGCG 420  
361 CGAACACGCTCGGCGACTCCATTCGGGGCTGTCTTACTCCCACTCTCGCGAGACTGGCG 420  
421 GCGGGCGAGCGAGGCCCAAGCTGGGAGCTCAGCTCGCGCGAGCCGAGCTGCCCTGTG 480  
421 GCGGGCGAGCGAGGCCCAAGCTGGGAGCTCAGCTCGCGCGAGCCGAGCTGCCCTGTG 480  
481 TGTCCCGCGCTCCCGGGCTTGGTGGCGCTCTGGACGCGCGGGGAGCGGACCAACGC 540  
481 TGTCCCGCGCTCCCGGGCTTGGTGGCGCTCTGGACGCGCGGGGAGCGGACCAACGC 540  
541 CGGAGGATGGACGAAGGTCTCGGACATTTGCGCGCGCGGGCGGGTGGCAGGGTGG 600  
541 CGGAGGATGGACGAAGGTCTCGGACATTTGCGCGCGCGGGCGGGTGGCAGGGTGG 600  
601 AAGCGAGGGCGTGGCGAGAGCTGCGAGCGGGCGAGAACCGCTGGGGGAACCCCTTG 660  
601 AAGCGAGGGCGTGGCGAGAGCTGCGAGCGGGCGAGAACCGCTGGGGGAACCCCTTG 660  
661 GTCGGCTCTCGCGTCTGAGATCCCGGAAAGAGCAGCGGCGCGGAAAGCGGCCAG 720  
661 GTCGGCTCTCGCGTCTGAGATCCCGGAAAGAGCAGCGGCGCGGAAAGCGGCCAG 720  
721 GCTGGGCGAGGATCTAGAAAGCTGCTGCGGAGGCTCCCTGCCCGCGGGCTGCTG 780  
721 GCTGGGCGAGGATCTAGAAAGCTGCTGCGGAGGCTCCCTGCCCGCGGGCTGCTG 780  
781 TCATGAGCTCGTGAGAGCTGCTTCCCGCGGGAACCTTCTCGAGGGTCAACGTCCA 840  
781 TCATGAGCTCGTGAGAGCTGCTTCCCGCGGAGCCCTTCTCGAGGGTCAACGTCCA 840  
841 GGCACCGGCGCTCGGACACCCACCCCGCGCGGACCTGCTGGGTGCCCTTTAAC 900  
841 GGCACCGGCGCTCGGACACCCACCCCGCGCGGACCTGCTGGGTGCCCTTTAAC 900  
901 CCGGGCGGTAGCTCGTTAAGTGGCAAGTGTCCGGTCCGGAAACAGCGGAAACCCCAAT 960  
901 CCGGGCGGTAGCTCGTTAAGTGGCAAGTGTCCGGTCCGGAAACAGCGGAAACCCCAAT 960  
961 CCGGCTGCCGACCTCTGACCCCGCGCCCAAGCGGACACGAGACTGGGCTCCGAGC 1020  
961 CCGGCTGCCGACCTCTGACCCCGCGCCCAAGCGGACACGAGACTGGGCTCCGAGC 1020  
1021 CGCAGCGCTGCGGGACACCGGTGCTGCGGAAACGAGGACCTTTGTAACGCGACGTG 1080  
1021 CGCAGCGCTGCGGGACACCGGTGCTGCGGAAACGAGGACCTTTGTAACGCGACGTG 1080  
1081 TTTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTCTGTAAGCTTGGCGCT 1140  
1081 TTTGCTCTTTTGAAGAAACAGAAATGTTTAACTGTCTGTAAGCTTGGCGCT 1140  
1141 AAAAGATGCTGGGTGATCTAGATGCTAGGATCAGTTTGTCTTCAATGTAATGGACAG 1200  
1141 AAAAGATGCTGGGTGATCTAGATGCTAGGATCAGTTTGTCTTCAATGTAATGGACAG 1200  
1201 CCCGGACTCGGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGCTGGTGGC 1260

1201 CCGGACTCCGTACGGCACTAGCAGGGGACTGAAGCGTCTTCAGGTACTGCTGGTGGC 1260  
1261 GGTGATGCGCTACAGGCCGATCAGACAGTTTTGTGTCTTCTGGAACTTGACACTGCACCA 1320  
1261 GGTGATGCGCTACAGGCCGATCAGACAGTTTTGTGTCTTCTGGAACTTGACACTGCACCA 1320  
1321 CCGTAATGCTGAACCTGACCAATATTTACAGATCAGACGCGATCATCTTCTTCAACATGA 1380  
1321 CCGTAATGCTGAACCTGACCAATATTTACAGATCAGACGCGATCATCTTCTTCAACATGA 1380  
1381 TTTTACACAGTTTGAATTAATATGTTGGAATAATGTTAGAAATCAAAATTTACCAATACCCAC 1440  
1381 TTTTACACAGTTTGAATTAATATGTTGGAATAATGTTAGAAATCAAAATTTACCAATACCCAC 1440  
1441 CTCAGGCTCTACTCTGTAATTTGAGCAGGTGTTTAACTCTTTTGTACCTCAGCTTCT 1500  
1441 CTCAGGCTCTACTCTGTAATTTGAGCAGGTGTTTAACTCTTTTGTACCTCAGCTTCT 1500  
1501 TCATTTACAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCCAGCACT 1560  
1501 TCATTTACAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCCAGCACT 1560  
1561 TGGGAGGCGGAGGCGAGCCGATCACTTCAGAGTTCAGAACTTTTCAGACCAAGCTTGGTCAAC 1620  
1561 TGGGAGGCGGAGGCGAGCCGATCACTTCAGAGTTCAGAACTTTTCAGACCAAGCTTGGTCAAC 1620  
1621 ATGGGTGAAACCCCGCTCTACCAAAAATATAAAACTTTAGCTGAGTGTGGTAGCGCATG 1680  
1621 ATGGGTGAAACCCCGCTCTACCAAAAATATAAAACTTTAGCTGAGTGTGGTAGCGCATG 1680  
1681 ACTGTAATCCCAGCAACTCAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
1681 ACTGTAATCCCAGCAACTCAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
1741 GGTTCAGTGCAGTGCAGTGCATCCACTGCACCTTCAGCTTTGGCGGACAGAGGAGACTCT 1800  
1741 GGTTCAGTGCAGTGCAGTGCATCCACTGCACCTTCAGCTTTGGCGGACAGAGGAGACTCT 1800  
1801 GCTTAAAAATATAAATAATTTTTTAAAAATAAGGGGTACTAATATCTACCTTTAAAG 1860  
1801 GCTTAAAAATATAAATAATTTTTTAAAAATAAGGGGTACTAATATCTACCTTTAAAG 1860  
1861 GATGAGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTAATCT 1920  
1861 GATGAGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTAATCT 1920  
1921 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTCAGCCAGGAGTTTGAGACTAGTCT 1980  
1921 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTCAGCCAGGAGTTTGAGACTAGTCT 1980  
1981 GGGCAACAGAGACATGCTCTATAGTTGTTGTTGTTGTTTAAACAGGTGTGTGTT 2040  
1981 GGGCAACAGAGACATGCTCTATAGTTGTTGTTGTTGTTTAAACAGGTGTGTGTT 2040  
2041 GTGCACTTCGAGTCCAGCTCTAGGGAGGCTGAGGTGGGAGACTGCTTGGAGCCGAGGA 2100  
2041 GTGCACTTCGAGTCCAGCTCTAGGGAGGCTGAGGTGGGAGACTGCTTGGAGCCGAGGA 2100  
2101 GGTTCAGGCTGCACTGAGCCATGATTTGTGCCACTGCACTCCAGCCCTGGGCAACACAGCAA 2160  
2101 GGTTCAGGCTGCACTGAGCCATGATTTGTGCCACTGCACTCCAGCCCTGGGCAACACAGCAA 2160  
2161 GACCTTGTCTCAAAAACAAAACAAAAGCATCTCATAAAGTGTCTGGCTCTTATATGAT 2220  
2161 GACCTTGTCTCAAAAACAAAACAAAAGCATCTCATAAAGTGTCTGGCTCTTATATGAT 2220  
2221 CAATATGCTGGTGGATCTTGAATCTTCTGACTCAGATCTCATAGATTTTCTGA 2280  
2221 CAATATGCTGGTGGATCTTGAATCTTCTGACTCAGATCTCATAGATTTTCTGA 2280  
2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAAAACCGTCAAGGCACTCCCTTCTGCC 2340

```

Db      2281  ACTTTTGGAGAACTCTTGCCCTCTCTGCAATTTGCAAAACCGTCAAGGACACTCCCTCTTGCC 2340
QY      2341  ACCACAAAGACATTTGATTTTAACTTGACTATGCTCTCTCTCTCTCAACTTTAGGTAAA 2400
Db      2341  ACCACAAAGACATTTGATTTTAACTTGACTATGCTCTCTCTCTCTCAACTTTAGGTAAA 2400
QY      2401  TTAATCTTGGTCAGGCTTCTCTGAACAGCCCTTTAGTCACTATGCAATGAATACATGCG 2460
Db      2401  TTAATCTTGGTCAGGCTTCTCTGAACAGCCCTTTAGTCACTATGCAATGAATACATGCG 2460
QY      2461  CCTACAGCT 2469
Db      2461  CCTACAGCT 2469

RESULT 2
US-10-829-118-2
; Sequence 2, Application US/10829118
; Publication No. US20040191759A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/10/829,118
; CURRENT FILING DATE: 2004-04-20
; PRIOR APPLICATION NUMBER: US/05/891,711
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-829-118-2

Query Match      100.0%; Score 2469; DB 19; Length 2469;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TTGAGTCTGTCTGGAGGCTCCGGCCAGACAGCGGGCGTATTGTTTCACTCGTGGTAATGCT 60
Db      1  TTGAGTCTGTCTGGAGGCTCCGGCCAGACAGCGGGCGTATTGTTTCACTCGTGGTAATGCT 60
QY      61  CATTTACGTTAAAGAAACAGGCAACGGAACAGCTGCCGGAGCGCGAGACCCCGCA 120
Db      61  CATTTACGTTAAAGAAACAGGCAACGGAACAGCTGCCGGAGCGCGAGACCCCGCA 120
QY      121  GGGCCGGGTACAGGCAAGCTGTGTCCAAACAAAGCGCGGAGCGCCCGCCACCTCCC 180
Db      121  GGGCCGGGTACAGGCAAGCTGTGTCCAAACAAAGCGCGGAGCGCCCGCCACCTCCC 180
QY      181  CCGACCCGGCCCGGCCCCCGCAGCCCTCGCTCGGGGCTCGGACGCAACCGGACACCT 240
Db      181  CCGACCCGGCCCGGCCCCCGCAGCCCTCGCTCGGGGCTCGGACGCAACCGGACACCT 240
QY      241  GAGCGAGCGGCGCCCAACCGCTAGCGGAGCGGGTCCGGGAGCGCGCGCGCGCGGCTG 300
Db      241  GAGCGAGCGGCGCCCAACCGCTAGCGGAGCGGGTCCGGGAGCGCGCGCGCGCGGCTG 300
QY      301  ACCTACTCTGCGCGCGGAGCTCAGGCGCGGCGCGCGCGGATGACGGCGCTCCGCGG 360
Db      301  ACCTACTCTGCGCGCGGAGCTCAGGCGCGGCGCGCGCGGATGACGGCGCTCCGCGG 360
QY      361  CGAACAGCTTGGGCACTCCATTTGGGGCTGTTTACTTCCCAACTCTCGGAGACTGGGCG 420
Db      361  CGAACAGCTTGGGCACTCCATTTGGGGCTGTTTACTTCCCAACTCTCGGAGACTGGGCG 420
QY      421  GCGGGGCGAGGAGGCCCAAGCTGAGGAGGCTCAGCTCCGCGAGCCAGCGTGCCTGTC 480
Db      421  GCGGGGCGAGGAGGCCCAAGCTGAGGAGGCTCAGCTCCGCGAGCCAGCGTGCCTGTC 480
QY      481  TGTCCCGCGCTCCCGGGCTTGGTGGCGCTCTGGAGCGCGTGGGCGAGCGGACCAAGC 540

```

```

Db      481  TGTCCCGCGCTCCCGGGGCTTGGGTGCGCGCTCTTGGAACCGCTGGGACGCGGACCAAGC 540
QY      541  CGGAGGATGGACGAAAGTGTCTCGGACATTTTCGGCGCGGCGGGCGCGGTGGCAGGTTGG 600
Db      541  CGGAGGATGGACGAAAGTGTCTCGGACATTTTCGGCGCGGCGGGCGCGGTGGCAGGTTGG 600
QY      601  AAGCGGAGGGGCGTGGCCAGCGAGCTGCGAGCGCGGAGAAACGCGCTGGGGGAAACCTTGG 660
Db      601  AAGCGGAGGGGCGTGGCCAGCGAGCTGCGAGCGCGGAGAAACGCGCTGGGGGAAACCTTGG 660
QY      661  GTCCGCTCTGCGCGCTCGCTTAGGATCCCGGAAAGGAGACGCGGCGCGGAAAGCGGCGAG 720
Db      661  GTCCGCTCTGCGCGCTCGCTTAGGATCCCGGAAAGGAGACGCGGCGCGGAAAGCGGCGAG 720
QY      721  GCTGGGCCAGGATCTAGAAAGACTGCTGCGCGAGGCTCCCTGCCCGCGGCGCTGCTG 780
Db      721  GCTGGGCCAGGATCTAGAAAGACTGCTGCGCGAGGCTCCCTGCCCGCGGCGCTGCTG 780
QY      781  TCATGGACTCGTGGAGAGCTGCTTCCCGCGGGAACCTTCTCTGCAAGGCTCCAGTCCA 840
Db      781  TCATGGACTCGTGGAGAGCTGCTTCCCGCGGGAACCTTCTCTGCAAGGCTCCAGTCCA 840
QY      841  GGCACCGGCGGCTCGGACACCCCGCGGCGGCGACCTGCGCTGGGTGCGGCTTAAAC 900
Db      841  GGCACCGGCGGCTCGGACACCCCGCGGCGGCGACCTGCGCTGGGTGCGGCTTAAAC 900
QY      901  CCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAACACCGGAAACCCCAAT 960
Db      901  CCGGCGGTAGCTCGTTAAGATGGCGAAGTGTCCGGTCCGGAACACCGGAAACCCCAAT 960
QY      961  CCGGCTGCGCGGCTCCTGACCCCGCGGCGGCGGAGACAGACTGGGCTCCCGAGC 1020
Db      961  CCGGCTGCGCGGCTCCTGACCCCGCGGCGGCGGAGACAGACTGGGCTCCCGAGC 1020
QY      1021  CGGAGCGGCTGCGGCGGACACCGGTCGGTGGGAAACGAGGAGACTTGTGAACGCGAGTG 1080
Db      1021  CGGAGCGGCTGCGGCGGACACCGGTCGGTGGGAAACGAGGAGACTTGTGAACGCGAGTG 1080
QY      1081  TTTGCTCTTTTGAAGAAACAGAAATAAATGCTTAACTGTCTGAAAGCTTGGCGCCT 1140
Db      1081  TTTGCTCTTTTGAAGAAACAGAAATAAATGCTTAACTGTCTGAAAGCTTGGCGCCT 1140
QY      1141  AAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGACAG 1200
Db      1141  AAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGACAG 1200
QY      1201  CCGGACTCCGTACGCGACTAGCAGGGGACTGAAAGCGTTCAGGTATGCTGGTGGG 1260
Db      1201  CCGGACTCCGTACGCGACTAGCAGGGGACTGAAAGCGTTCAGGTATGCTGGTGGG 1260
QY      1261  GGTGATGCGGTACAGCGCGATCAGACAGTTTGTGCTCTTGGAACTTGACACTGACCA 1320
Db      1261  GGTGATGCGGTACAGCGCGATCAGACAGTTTGTGCTCTTGGAACTTGACACTGACCA 1320
QY      1321  CGGTAATGCTGAATCAGCAATATTTACAGATCAGACGATCATCTTCTTCAACATGA 1380
Db      1321  CGGTAATGCTGAATCAGCAATATTTACAGATCAGACGATCATCTTCTTCAACATGA 1380
QY      1381  TTTAAACACAGTTGACTTAATATATGTTGGATAAATGTAAGATCACAAAATACCAACCC 1440
Db      1381  TTTAAACACAGTTGACTTAATATATGTTGGATAAATGTAAGATCACAAAATACCAACCC 1440
QY      1441  CTGAGGCTTCTACTTCTGTAATTTTGGAGGCTGTTTAACTCTTGTGTAATGCTGCTTCT 1500
Db      1441  CTGAGGCTTCTACTTCTGTAATTTTGGAGGCTGTTTAACTCTTGTGTAATGCTGCTTCT 1500
QY      1501  TCAATTAACAAATAGGGGACTAGCGGCGGGGTGCTCGCGCTGTAAATCCAGACT 1560
Db      1501  TCAATTAACAAATAGGGGACTAGCGGCGGGGTGCTCGCGCTGTAAATCCAGACT 1560
QY      1561  TGGGAGGCGCGAGCGGCGGATCACTTGAAGGTGAGAAATTTTCAAGACGAGCTGGTCAAC 1620
Db      1561  TGGGAGGCGCGAGCGGCGGATCACTTGAAGGTGAGAAATTTTCAAGACGAGCTGGTCAAC 1620

```



```

QY 1621 ATGGGTGAACGGCGGCTCTACCAAAATATAAAATCTTAGCTGAGTGGTAGCGCATG 1680
Db 1621 ATGGGTGAACGGCGGCTCTACCAAAATATAAAATCTTAGCTGAGTGGTAGCGCATG 1680
QY 1681 ACTGTAATCCAGCAACTCAGGAGGCTGAGGAGAGAGAAATCGCTTGAACCTTGGGAGCGGA 1740
Db 1681 ACTGTAATCCAGCAACTCAGGAGGCTGAGGAGAGAGAAATCGCTTGAACCTTGGGAGCGGA 1740
QY 1741 GGTTCAGTGAAGTGAAGTGAACCTACCTGACCTCCAGCTTGGGCGGACAGAGGAGACTCT 1800
Db 1741 GGTTCAGTGAAGTGAAGTGAACCTACCTGACCTCCAGCTTGGGCGGACAGAGGAGACTCT 1800
QY 1801 GCCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
Db 1801 GCCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
QY 1861 GATGAGGTTAAATTAAGTACACATTAAGCCCTAGCGAGTGGCTTATGCGCTGTAACT 1920
Db 1861 GATGAGGTTAAATTAAGTACACATTAAGCCCTAGCGAGTGGCTTATGCGCTGTAACT 1920
QY 1921 CAACACTTTGGAGTCTGTGGCGGAGGATCACTTGAAGCCAGGAGTTGAGACTAGTCT 1980
Db 1921 CAACACTTTGGAGTCTGTGGCGGAGGATCACTTGAAGCCAGGAGTTGAGACTAGTCT 1980
QY 1981 GGGCAACAGAGACATCTCTATAGTTGTGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGT 2040
Db 1981 GGGCAACAGAGACATCTCTATAGTTGTGTTTGGTTTGGTTTGGTTTGGTTTGGT 2040
QY 2041 GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCGCTGAGCCACAGGA 2100
Db 2041 GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGGTGGAGGACTGCGCTGAGCCACAGGA 2100
QY 2101 GGTGAGGCTGAGTGAAGCATGATTTGTGCACTGCACTGCACTGCACTGCACTGCACTGCA 2160
Db 2101 GGTGAGGCTGAGTGAAGCATGATTTGTGCACTGCACTGCACTGCACTGCACTGCA 2160
QY 2161 GACCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2220
Db 2161 GACCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2220
QY 2221 CAATATGTGGTGGTGAATCTTCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
Db 2221 CAATATGTGGTGGTGAATCTTCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
QY 2281 ACTTTTGGAGAACTCTTGCCTCTCTGCAATTTGCAACCGGTCAAGGCACTCCCTTCTGCC 2340
Db 2281 ACTTTTGGAGAACTCTTGCCTCTCTGCAATTTGCAACCGGTCAAGGCACTCCCTTCTGCC 2340
QY 2341 ACCACAAAGCAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGA 2400
Db 2341 ACCACAAAGCAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGAATTTGA 2400
QY 2401 TTAATCTTGTGAGGTTCTCTCAACAGCCCTTTAGTCACTATGCAATTTGAATTTGAATTT 2460
Db 2401 TTAATCTTGTGAGGTTCTCTCAACAGCCCTTTAGTCACTATGCAATTTGAATTTGA 2460
QY 2461 CCTACAGCT 2469
Db 2461 CCTACAGCT 2469

```

```

RESULT 3
US-09-891-711-1
; Sequence 1, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 980
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-891-711-1

```

```

Query Match 39.7%; Score 980; DB 9; Length 980;
Best Local Similarity 100.0%; Pred. No. 6e-255;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 CTCCGCGGCGAACAAGCGCTTGGGCACTCCATTCTGGGGCTGTTTACTCCCAACTCTCCGCGAG 412
Db 1 CTCCGCGGCGAACAAGCGCTTGGGCACTCCATTCTGGGGCTGTTTACTCCCAACTCTCCGCGAG 60
QY 413 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 472
Db 61 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 473 GCCCTGTCTGCCCGGCTCCCGGGGCTTGGCGGCGCTCTGGACGCGCTGGGCGAGCGG 532
Db 121 GCCCTGTCTGCCCGGCTCCCGGGGCTTGGCGGCGCTCTGGACGCGCTGGGCGAGCGG 180
QY 533 GACCAACGCGGCGGAGGATGGAACGAGTGTCTCGCGACATTTGCGGCGGCGGCGGCGGCGG 592
Db 181 GACCAACGCGGCGGAGGATGGAACGAGTGTCTCGCGACATTTGCGGCGGCGGCGGCGGCGG 240
QY 593 CAGGCTGGAAGCGGAGGCGGCTGGCGAGGAGCTGCGAGGCGGCGGAGAACGCGCTGGGGG 652
Db 241 CAGGCTGGAAGCGGAGGCGGCTGGCGAGGAGCTGCGAGGCGGCGGAGAACGCGCTGGGGG 300
QY 653 AACCTTGTCTCGCTCTGCGGCTCTGAGTATCCCGAAGGAGGACGCGGCGGCGGAA 712
Db 301 AACCTTGTCTCGCTCTGCGGCTCTGAGTATCCCGAAGGAGGACGCGGCGGCGGAA 360
QY 713 GCGGCGGAGGCTGCGGCGGAGGATCTAGAAAGCTGCGCTGCGCGCAGGCTCCCTGCGCGCGG 772
Db 361 GCGGCGGAGGCTGCGGCGGAGGATCTAGAAAGCTGCGCTGCGCGCAGGCTCCCTGCGCGCGG 420
QY 773 GCGTGTGTATGAGTCTGTGAGAGTCTGCTTCCGCGCGGACCTTCTTCTGAGGGGTC 832
Db 421 GCGTGTGTATGAGTCTGTGAGAGTCTGCTTCCGCGCGGACCTTCTTCTGAGGGGTC 480
QY 833 CAGTCCAGGCGACCGGCGGCTCGGACACCCACCCCGGCGGCGGCGGCGGCGGCGGCGGTC 892
Db 481 CAGTCCAGGCGACCGGCGGCTCGGACACCCACCCCGGCGGCGGCGGCGGCGGCGGCGGTC 540
QY 893 CCTTAAACCGGCGGCTAGCTCGTTAAGATGCGAAGTGTCCGGTCCGGAAACACGCGGAA 952
Db 541 CCTTAAACCGGCGGCTAGCTCGTTAAGATGCGAAGTGTCCGGTCCGGAAACACGCGGAA 600
QY 953 CCGCAAAATCCCGCTGCGCGGACCTCTGACCCCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1012
Db 601 CCGCAAAATCCCGCTGCGCGGACCTCTGACCCCGCGGCGGCGGCGGCGGCGGCGGCGGCG 660
QY 1013 TCCGACGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1072
Db 661 TCCGACGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
QY 1073 GCCACGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 1132
Db 721 GCCACGTGTTGCTCTTTTGAAGAAACAAAGATTAATGTTTAAACTGTCTGAAAAGCT 780
QY 1133 TCGCGCTTAAAGATGCTGGGTGACTTAGATGCTAGGATCACTTGTGTTTCAATGATAA 1192
Db 781 TCGCGCTTAAAGATGCTGGGTGACTTAGATGCTAGGATCACTTGTGTTTCAATGATAA 840
QY 1193 TGGACCGGCGGCGGCTGCGTACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1252
Db 841 TGGACCGGCGGCGGCTGCGTACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
QY 1253 TGGTGGGCGGCTGATGCGCTTACAGGCGGATCAGACAGTGTGTTGCTCTTCTGGAACCTTGACA 1312

```





QY 1867 GGTAAATTA-----AGTACACACATAAGCCCTAGCGCAGTGGCTTATGCTGTAA 1917  
Db 259065 GCAAGCATTAATAATAATAATAATAATAAGCCAGGCACAGTGGCTCATAGCTGTAA 259006  
QY 1918 TCTCAACACCTTTGGGAGTCTGTGCGCGGAGGATCACTTGAGCCAGGAGTTTGAGACTAG 1977  
Db 259005 TCCAGCACCTTCGGGAGGCGGAGGAGGAGGATCACTTGAGGTCAGGGGTTGAGACAG 258946  
QY 1978 TCTGGGCAACAGACAGATCTCTATAGTTGTGTTTGGTTTGTGTTTATACAGGTGTGT 2037  
Db 258945 CTTGGGCAACATGGTGAACCTGTCTCTACAAAAATAAAAAATTTAGCCAGCGTGTGT 258886  
QY 2038 GGTGTGCACCTGCAGTCCAGTCCAGTACTAGGAGGCTGAGGTGGGAGGACTGCCTGAGCCCA 2097  
Db 258885 GGTGCATGGCTATATCCAGCTACTTTGGGAGACTGAGGAGAGAAATCACTTGAACCCG 258826  
QY 2098 GAGGTTCGAGGCTGCAGTGCAGTGCATGATTTGTGCACTTGCACTCCAGCTGGGCAACACAG 2157  
Db 258825 AGAGATGAAGTTGCAGTGAGCTGAGATTGCGGCACTGCACTCCAGCTGGCCACAGAG 258766  
QY 2158 CAAGACCTTGTCTCAAAAAACAAACAAA 2185  
Db 258765 TGAGACCTTGTCAAAAAAATAAAAAA 258738

## RESULT 6

US-10-322-281-170/c  
; Sequence 170, Application US/10322281  
; Publication No. US20040126762A1

; GENERAL INFORMATION:

; APPLICANT: David W. Morris

; APPLICANT: Marc S. Malandro

; TITLE OF INVENTION: Novel Compositions and Methods in Cancer

; FILE REFERENCE: 529452001000

; CURRENT APPLICATION NUMBER: US/10/322,281

; CURRENT FILING DATE: 2002-12-17

; NUMBER OF SEQ ID NOS: 866

; SOFTWARE: PaeCSQ for Windows Version 4.0

; SEQ ID NO 170

; LENGTH: 100445

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(100445)

; OTHER INFORMATION: n = A,T,C or G

US-10-322-281-170

Query Match 13.4%; Score 330.8; DB 19; Length 100445;  
Best Local Similarity 72.3%; Pred. No. 5.9e-78;  
Matches 506; Conservative 0; Mismatches 172; Indels 22; Gaps 5;

QY 1494 AGCTTCTTCATTACAAAAATAGGGGTACTAGCCAGCGGGTGGCTCGCGCCCTGTAAATCC 1553  
Db 47292 AGATTCTGTCTTAAATAATAAAGATTAACGGCCGGCGGGTGTCTATGCTGTAAATCC 47233  
QY 1554 CAGCCTTGGGAGGCGGAGCGGAGCCCGATCACTTGAGGTTCAGAGTTTTCAGACCAAGCCT 1613  
Db 47232 CAGCCTTGGGAGGCGGAGGAGGTGGATCAC - GAGGTTCAGAGATTGAGACCATCTCT 47175  
QY 1614 GGTCAACATGGGTGAACCGCGCTCTACCAAAAAATATAAAAAATTTAGCTAGTGTGGTA 1673  
Db 47174 GGTAAACAT - GGTGAACCCCGTCTCTACTAAAAATAACAAAAATTTAGCTGGCGGTGGT 47116  
QY 1674 GCGCATGACTGTAATCCAGCAACTCAGAGGCTGAGGCA - GAGAATCGCTTCAACTGG 1732  
Db 47115 GCAGGTGCTTGTAGTCTCAGCTACTCAGGAGGCTGAGGAGGAGATGGGTGAACCCGG 47056  
QY 1733 GAGCGGAGGTTGAGTGAAGTGAATCGTACCACTGCACTCCAGTCTGGGCGACAGAGC 1792  
Db 47055 GAGCGGAGGTTGAGTGAAGTGAATCGTGCACACTGCACTCCAGCTGGGCGACAGAGC 46996

QY 1793 GAGACTCTGCCCTTAAAAATAATAATAATAATTTTAAAAAAAATAGGGGTACTAATATCTA 1852  
Db 46995 AAGACTCCGCTCTCAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATA 46952  
QY 1853 CTTAAAGGATGAGGGTTAAATTTAAAGTACACATAAAGCCCTTAGCGCAGTGGCTTATGCC 1912  
Db 46951 ATAAAAAAGATAAAAAACAAAAACAAAAACAAATCAGGCGCGGCGGTGGCTCACACC 46892  
QY 1913 TGTATCTCAACACCTTTGGGAGTCTGTGCGGAGGATCACTTGAGCCCGGAGGTTTGGAG 1972  
Db 46891 TGTAAATCCAGAACTTTGGGAGGCTGATGGGCGCGATCACTTAAAGTTCAGAGTTTCGAG 46832  
QY 1973 ACTAGTCTGGGCAACA - GAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTACCAAG 2030  
Db 46831 ACAGCTTGGCAACATGTTGAAACCTTATCTCAATAATAAATAACAAAAAATTAGCCAG 46772  
QY 2031 GTGTGGTGGTGTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGTGGGAGGACTGCCT 2090  
Db 46771 ATGTGTGATGCGCACCTCTGTAATCCAGATACTCAGAGGCTGAGGACCAAGATCGCTC 46712  
QY 2091 GAGCCAGGAGTGCAGGCTGCAGTGCAGGATGATTTGCCACTGCAGCTCCAGCCTGGGC 2150  
Db 46711 GTACCCAGGAGCAGAGGTTGCAGTGAGCAAGATCTAGCCACTGCAGCTCCAGCCTGGAT 46652  
QY 2151 AACACAGCAAGACCTTGTCTCAAAAAACAAACAAAAAGCAT 2190  
Db 46651 GACAAAGTGAGACTTGTCTCAAAAAAATAAAAAAATAAAAAA 46612

## RESULT 7

US-10-723-860-2768/c

; Sequence 2768, Application US/10723860

; Publication No. US20040253606A1

; GENERAL INFORMATION:

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsburg, Wendy M.

; APPLICANT: Zlotnik, Albert

; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &

; FILE REFERENCE: 05882.0193.NPUS01

; CURRENT APPLICATION NUMBER: US/10/723,860

; CURRENT FILING DATE: 2003-11-26

; PRIOR APPLICATION NUMBER: 60/429,739

; PRIOR FILING DATE: 2002-11-26

; NUMBER OF SEQ ID NOS: 8393

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2768

; LENGTH: 147620

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-723-860-2768

Query Match 13.3%; Score 327.4; DB 20; Length 147620;

Best Local Similarity 71.0%; Pred. No. 5.9e-77;

Matches 489; Conservative 0; Mismatches 196; Indels 4; Gaps 4;

QY 1504 TTACAAAAATAGGGGTACTAGCCAGCGGGTGGCTCGCGCCCTGTAAATCCAGCATTGG 1563  
Db 107541 TCAAAAAATAATAAATAGGCGGGTGTGTGGCTCACGCTGTAAATCTCAGCATTGG 107482  
QY 1564 GGAGCGGAGGCGGAGGATCACTTGAGGTTCAGAGTTTTCAGACCAAGCTGGTCAACATG 1623  
Db 107481 GGAGCGGAGGCGGCGGATCACTTGAGGTTCAGAGTTTCAGACCAAGCTGGCCAACT- 107423  
QY 1624 GGTGAACCGCGCTCTACCAAAAAATATAAAAACTTAGCTGAGTGTGGTAGCGCATGACT 1683  
Db 107422 GGTGAACCGCGCTCTCTACTAAAAATAACAAAA - TTAGCTGGCGGTGGTGGTGGCT 107364  
QY 1684 GTAATCCAGCAACTCAGAGGCTGAGGCA - GAGATCCCTTGAACCTGGGCGGAGG 1742  
Db 107363 CTAAATCCAGCTACTTTGGAGGCTGAGGCAAGAAATCACTGAATCCAGGAGGCGGAGG 107304  
QY 1743 TTGCAGTGAGCTGAGATCGTACCACTGCACTCCAGCTTGGGCGACAGAGCGAGACTCTGC 1802





```
; LENGTH: 354592
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-70

Query Match      13.1%; Score 323.6; DB 22; Length 354592;
Best Local Similarity 71.0%; Pred. No. 9.4e-76;
Matches 487; Conservative 0; Mismatches 189; Indels 10; Gaps 4;

QY 1508 AAAAATAGGGGTACTAGCAGCGGGGTGGCTCGCGCTGTATCCACAGCTTGGGGAG 1567
Db 91139 AAGAATAGGAATTTGCGCGCGGGGTGGCTCACGCTGTATATCCACAGCTTGGGGAG 91080
QY 1568 GCGGAGCAGCGGATCACTTGAGGTGAGAAGTTTCAGACCGACCTGGTCAACATGGGTG 1627
Db 91079 GCGGAGCGGGCGGATCAC--GAGGTGAGAGATCGAGACCATCCCGG--CTAAACGGGTG 91023
QY 1628 AAACGCGGCTCTACCAAAATATAAAAACTTAGCTGAGTGTGGTAGCGCATGACTGTAA 1687
Db 91022 AAACCCGCTCTACTTAAAAATACAAAAATTAGCGGGGTAGTGGCGGGCGCTGTAG 90963
QY 1688 TCCAGCACTCAGGAGGCTGAGGCA-GAGAATCGCTTGAACCTGGAGCGCGAGTTGC 1746
Db 90962 TCCAGCTACTTGGGAGGCTGAGCGAGGAGATGGCGTGAACCCGGAGCGGAGCTTCG 90903
QY 1747 AGTGAGCTGAGATCGTACCACCTGCCTCCAGCTTGGCGGACAGAGGAGACTCTGCCTTA 1806
Db 90902 AGTGAGCGGAGATTTGGCGCACTGCATCTCGAGCTTGGCGGACAGAGGAGACTCCGCTCA 90843
QY 1807 AAAATAAATAAATAAATTTTTTAAAAAANA- - - - -GGGCTACTAATATCTTACCTTAAAG 1860
Db 90842 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 90783
QY 1861 GATGAGGTTAAATTAAGTACACATATAGCCCTAGCGGAGTGGCTTATGCTGTAACTCT 1920
Db 90782 GACAAACGACAGCAAAATTAAGAATATTAGCGGAGTGGCGTAGCTCATGCTGTAAATGC 90723
QY 1921 CAACACTTTGGGAGTCTGGCGGAGGATCACTTGAGCCCGAGGATTTGAGACTAGTCT 1980
Db 90722 CAGCACTTTGGGAGCGCGAGCGGATGATCACTGAGGCGAGGATTTGAGACACAGCT 90663
QY 1981 GGGCAACAGACATGCTCTATAGTTGTGTTGGTTTTTTTACCAAGGTGTGTGTG 2040
Db 90662 GGCACATGCTGAAACCCGCTCTCTACTAAAAATACAAAAATAGCTGGGCATTTGGCG 90603
QY 2041 GTCCACTGCGAGTCCAGCTACTAGGAGGCTGAGGTGGAGGAGTGCCTGAGCCAGGA 2100
Db 90602 ACGGCTGTAGTCCAGCTACTCAAGAGGCTGAGGCGAGGAGATCACTTAAACCCAGGA 90543
QY 2101 GGTGAGGCTGCAATGAGCCATGATTTGCACTGCACTCCAGCTGGGCAACACAGCAA 2160
Db 90542 GATGGAGGTTGCGGTGAGCGAGGATCATGCTCTGCACTCCAGCTGGGCGAGTGAACAA 90483
QY 2161 GACCTGCTCAAAAACAAACAAAAA 2186
Db 90482 GACTCCATCTCAAAAATAAATAA 90457

RESULT 12
US-10-719-993-6815/c
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6815
; LENGTH: 1980090
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)....(1980090)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6815

Query Match      13.1%; Score 322.6; DB 20; Length 1980090;
Best Local Similarity 71.6%; Pred. No. 3.8e-75;
Matches 489; Conservative 4; Mismatches 178; Indels 12; Gaps 5;

QY 1506 ACAAATAAGGGTACTAGCAGGGGGGTGGCTCGCGCTGTATATCCACAGCTTGGGG 1565
Db 1405235 ACAGATTATACAGATGACGCGCGGCGACGCTGGCTCACGCTGTATATCCACAGCTTGGC 1405176
QY 1566 AGGCGGAGCGCGGATCACTTGAGGTGAGAGTTTCAGACGAGCTGTGTCACATGGG 1625
Db 1405175 AGGCGGAGCGCGGATCAC--GAGGTGAGAGATTCAGACCATCTGCTTAACA-AG 1405119
QY 1626 TGAACGCGCGCTCTACCAAAATATAAAAACTTAGCTGAGTGTGGTAGCGCATGACTGT 1685
Db 1405118 AGAAACYCGCTCTCTACTAAATATACAAAAATTAGCYRGGCTGGGGCGGCGCTCT 1405059
QY 1686 AATCCAGCACTCAGGAGGCTGAGGCA-GAGAAATCGCTTGAACCTGGAGCGGAGGTT 1744
Db 1405058 AGTCCAGCTACTCGGAGGCTGAGCGAGGAGATGGCGTGAACCCGGGAGCGGAGCTT 1404999
QY 1745 GCAGTGAGCTGAGATCGTACCACTGCCTCCAGCTTGGCGGACAGAGCGAGACTCTGCCT 1804
Db 1404998 GCAGTGAGCGGAGATGGCACCTGCCTCCAGCTTGGCGGACAGAGCGAGACTCTGTCT 1404939
QY 1805 TAAAAATAAATAAATAAATTTTAAAAAATAAATAGGGGTACTAATATCTACTTTAAAGGATG 1864
Db 1404938 CAAAAAATAAATAAATAAATGATTACAGATGACAATGTGGGATGCAGCTAC-----GTAG 1404886
QY 1865 AGGGTTAAATTAAGTACACATATAAGCCCTAGCGAGTGGCTTATGCTGTATCTCAAC 1924
Db 1404885 TGTGTAGAAAGAAATGATAAACTTTGGCGGCGGAGTGGCTCAVCCCGTATATCCCAAC 1404826
QY 1925 ACTTTGGGAGTCTGGCGGAGGATCACTTGAGCCCGAGGATTTGAGACTAGTCTGGGC 1984
Db 1404825 ATTCTGGGAGGCTGAGGAGGCGGATCACTGAGGTGAGGATTTTCCAGACCGCTTAGCC 1404766
QY 1985 AACA-GAGACATGCTCTATAGTTGTGTTGGTTTTTTTACCAAGGTGTGTGTGTG 2043
Db 1404765 AACATGGCAAAACCTCTCTACTAAAAATAAATAAATAAATGGCTCGGTGTGTGTGCGAGG 1404706
QY 2044 CACCTGAGTCCAGCTACTAGGAGGCTGAGGTGGGAGTGTGCTGAGCCAGGAGGT 2103
Db 1404705 TGCTGTATATCCAGCTGTCTCGGAGGCTGAGGAGGAGATCACTTGAACCCAGGAGGC 1404646
QY 2104 CGAGGTGAGTGGAGCCATGATTTGCCACTGCACTCCAGCTTGGGCAACACAGCAAGAC 2163
Db 1404645 GGAGGTTGCAGTGAGCGGAGATAGCCCACTGCCTCCAGCTTGGGCGAGAGCGAGAC 1404586
QY 2164 CTGTCTCAAAAACAAACAAAAA 2186
Db 1404585 TCGGTCTCAAAAACAAACAAAAA 1404563

RESULT 13
US-10-741-600-17676/c
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
```

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 17676  
LENGTH: 1980090  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(1980090)  
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-5)  
US-10-741-600-17676

Query Match 13.1%; Score 322.6; DB 21; Length 1980090;  
Best Local Similarity 71.6%; Pred. No. 3.8e-75;  
Matches 489; Conservative 4; Mismatches 178; Indels 12; Gaps 5;

QY	1506	ACAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAATCCAGCACTTGGGG	1565
DB	1405235	ACAGTATTACAGATGACGGCGGCGACGGTGGCTCAGCGCTGTAATCCAGCACTTGGC	1405176
QY	1566	AGCCGAGGCGCGGATCACTTGGGTCAAGAGTTTCAGACAGCCTGTCAACATGGG	1625
DB	1405175	AGCCGAGGCGGCGGATCAC--GAGGTCAAGAGATCGAGACCATATGGCTAAACA-AGG	1405119
QY	1626	TGAACCGCGGCTCTACCAAAAAATATAAAAACTTAGTGTAGTGGTAGCGCATGACTGT	1685
DB	1405118	AGAAACVCCGTCTCTACTAAATAACAAAAATTAGYRGGCTGGGGGCGGCGCTGT	1405059
QY	1686	AATCCAGCAACTCAGAGGCTGAGGCA-GAGAAATCGTTGAACCTGGGAGGCGGAGTT	1744
DB	1405058	AGTCCAGCTACTTCGGAGGCTGAGGAGGAGTAATGGCGTGAACCGCGGAGGCGGAGCTT	1404999
QY	1745	GCAGTCAGCTGAGATCGTACCACTGCATCCAGCTTGGGCGGAGAGAGAGCTCTGCCT	1804
DB	1404998	GCAGTCAGCGGAGATGCACTGCACTCCAGCTTGGGCGGAGAGAGAGCTCTGTCT	1404939
QY	1805	TAAAAATAAATAAATAATTTTAAAAAATAGGGGTACTAATATCTACCTTAAAGGATG	1864
DB	1404938	CAAAAAAATAAATAAGTATTACAGATGACAACTGTGGGATGCAGCTAC-----GTAG	1404886
QY	1865	AGGGTAAATTAAGTACACACATAAGCCCTAGGCGAGTGCTTATGCTGTGTAATCTCAAC	1924
DB	1404885	TGTGTAGAAGAAATGTATAAATCTTGGCGGGCGAGTGGCTCAYGCCGTGTAATCCCAAC	1404826
QY	1925	ACTTTCGGAGTCGTGGCGGAGGATCACTTGAGCCGACGAGTTTCAGACTAGCTGGGC	1984
DB	1404825	ATTCTGGAGGCTGAGAGGCGGATCACTTGAGGTGAGGTTTCAGACCACTTGGC	1404766
QY	1985	AACA-GAGACATGCTCTATAGTGTGTGTTTGTGTTTGTGTTTACCAGGTGTGGTGTG	2043
DB	1404765	AACATGGCGAAACCTGTCTCTACTAAAAAATGCAAAAAATGGCTCGGTGTGGCAGG	1404706
QY	2044	CACCTGAGTCCAGCTACTAGGAGGCTGAGTGGAGAGCTGCGCTGAGCCAGAGGT	2103
DB	1404705	TGCGCTGTAATCCAGCTGCTCGGAGGCTGAGGAGGAGTAATCACTTGAACCCAGGAGG	1404646
QY	2104	CGAGGTGCACTGAGCATGATGTGCCACTGCACTCCAGCTTGGGCGGCAACACAGCAAGAC	2163
DB	1404645	GGAGGTGCACTGAGCGAGATAGCCCCACTGCACTCCAGCTTGGGCGGAGAGGAGAC	1404586
QY	2164	CTTGTCTCAAAAAACAAAAA 2186	
DB	1404585	TCGGTCTCAAAACAAAAA 1404563	

RESULT 14  
US-10-427-923-3/c  
Sequence 3, Application US/10427923  
Publication No. US2003020731A1  
GENERAL INFORMATION:  
APPLICANT: NEELAM, Beena et al.  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
THEREOF

FILE REFERENCE: CL001365  
CURRENT APPLICATION NUMBER: US/10/427,923  
CURRENT FILING DATE: 2003-05-02  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 105413  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(105413)  
OTHER INFORMATION: n = A,T,C or G  
US-10-427-923-3

Query Match	13.0%;	Score 322.2;	DB 17;	Length 105413;
Best Local Similarity	70.9%;	Pred. No. 1.3e-75;		
Matches	487;	Conservative 0;	Mismatches 188;	Indels 12; Gaps 4;
QY	1520	ACTAGCCAGCGGGTGGCTCGCGCTGTAATCCAGCACTTGGGAGGCGGAGGCGGC	1579	
DB	64357	AATGGCGGCGACGGTGGCTCAGCGCTGTAATCCAGCACTTGGGAGGCGGAGGCGGC	64298	
QY	1580	GGATCACTTTGAGGTCAAGAGTTTTCAGACCCAGCCTGTCTCAACATGGGTGAAACCGCGGCTC	1639	
DB	64297	GGATCAC--GAGGTCAAGAGATCGAGACCATCTTGGCTAACA-CAGTGAACCCCGCTC	64241	
QY	1640	TACCAAAAATATAAACTTAGTGTAGTGGTAGCGCATGACTGTAAATCCAGCAACTC	1699	
DB	64240	TACTAAAAACAAAAAATTAGCCGGCGTGGTGGTGGCGCTGTAGTCCAGCTACTT	64181	
QY	1700	AGGAGGCTCAGCA-GAGAAATCGTTGAACCTGGGAGGCGGAGTTTGCAGTGCAGT	1758	
DB	64180	GGAGGCTGAGGAGGAGTAATGGGTGAACCGGAGGAGAGCTTGCAGTGCAGCGCA	64121	
QY	1759	TGCTACCTGCACTCCAGCTTGGGCGACAGAGCCAGACTCTCTCCCTTAAAAATAAATAA	1818	
DB	64120	TGCACTGCACTCCAGCTTGGGCGACAGAGCCAGACTCTCTTCCAAAAAATAAATAA	64061	
QY	1819	TAAATTTTAAAAAATAGGGGTACTAATATCTACCTTAAAGATGAGGGTTAAATTAAG	1878	
DB	64060	AAAAAAGAGAGAAAAATGAGAAACAGACTTCTTTC-----TGGGATATGTTCAAAA	64009	
QY	1879	TACACATAAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCAACACTTTGGGAGTCTG	1938	
DB	64008	ATGTTATTCAAGCGCGGCGACCGTGGCTCAGCGCTGTAAATCCCATGCTTTGGAAGTGC	63949	
QY	1939	TGGCGGAGGATCACTTGAAGCCAGGAGTTTGAGACTAGTCTGGGCAACAGAGACATGTC	1998	
DB	63948	CGGAGGTGGATCACTTGAAGCCAGCTGTTCAAGACCAAGCTAGGCAACATGCTGAAC	63889	
QY	1999	TCTATAGTGTGTTGTTGTTTGTGTTTACAGGTGTGGTGTGACCTGCAAGTCCCAG	2058	
DB	63888	AGGTCTCTACTAAAAATACAAAAATTTAGCCAGGTGTGGTGTGACCTGCAATCCCAG	63829	
QY	2059	CTACTAGGAGGCTGAGGTGGGAGGACTCCCTGAGCCAGGAGGTGCGAGTGCAGTGAG	2118	
DB	63828	CTACTTGGAGGCTGAAATGGGAGGATCGCTTGAAGCCAGGAGGCCAAGTTGCACTGAG	63769	
QY	2119	CCATGATTTGCGCACTGCACTCCAGCCTGGGCAACACAGCAAGACCTTCTCTCAAAAA	2178	
DB	63768	CCAAGATCCCACTGCACTCCAGCCTGAGTGACAGAGTAAGACCTGTCTTTAAAAA	63709	
QY	2179	AACAAAAAGCATCTCATTAAGTGCTC 2205		
DB	63708	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	63682	

RESULT 15  
US-10-087-192-844/c  
Sequence 844, Application US/10087192  
Publication No. US20020182586A1  
GENERAL INFORMATION:  
APPLICANT: NEELAM, Beena et al.  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
THEREOF



Db 68222 AAAATAAAGGTTTTTTTAAA 68203  
Search completed: August 12, 2005, 08:55:00  
Job time : 1574 secs

APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: CANCER  
FILE REFERENCE: 529452000122  
CURRENT APPLICATION NUMBER: US/10/087,192  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 2059  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 844  
LENGTH: 91760  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(91760)  
OTHER INFORMATION: n = A, T, C or G  
US-10-087-192-844

Query Match		13.0%	Score 321.6;	DB 13;	Length 91760;
Best Local Similarity		72.1%	Pred. No. 1.8e-75;		
Matches 490;		Conservative 0;	Mismatches 179;	Indels 11;	Gaps 5;
Qy	1523	AGCCAGCGGGTGGCTGCGGCTGTAAATCCAGCACTTGGGAGGCGGAGGCGGCGGA	1582		
Db	68876	AGCCGGCGGAGTGGCTCAGCGCTGTAAATCCAGCACTTGGGAGGCGGAGGCGGAGTGA	68817		
Qy	1583	TCACCTTGAGGTCAGAGTTTCAGACCAGCCTGGTCAACATGGGTGAAACCCGCTCTAC	1642		
Db	68816	TCAC--GAGGTCAGCAGTTTGAGACCAGCCTGGTCAACAT--GGTGAACCCCGTCTCTAC	68760		
Qy	1643	CAAAAATATAAAATCTTAGCTGAGTGTGGTAGCGCATGACTGTAAATCCAGCAACTCAGG	1702		
Db	68759	TAAATACAAAATATTAGCCAG---GCACGGTGGCAGCGCTTAACCCCAACTACTCCGG	68703		
Qy	1703	AGGCTGAGGC--AGAGATCGGTTGAACCTGGGAGGCGGAGGTTGACGTGAGCTGAGATCG	1761		
Db	68702	AGGCTGATGCGGGAGAATCACTTGAACCGGAGGAGGTGAAGGTTGACGTGAGCCAGACCG	68643		
Qy	1762	TACCACCTGCACCTCCAGCTTGGGCGACAGAGGAGACTCTGCCTTAAAAATAAATAATAA	1821		
Db	68642	TGCCATTGCACCTCCAGCTTGGGCGACAGAGGAGACTCTGTCTCAAAAAAATAAATAA	68583		
Qy	1822	TTTTTAAAAAATAGGGGTACTAATATCTAACCTTAAAGGATGAGGGTTAAATTAAGTAC	1881		
Db	68582	AAAAAATAAAGCAAGGTCTGNAACAGTGACATAGTATGCCACGTTTGTGTTTTAAGA	68523		
Qy	1882	ACACATAGCCCTAGCGCAGTGGCTTATGCTGTAACTCAACACTTTGGGAGTCTGTGG	1941		
Db	68522	AAAGTGTAGGCTGGATGCAATGGCTCGCACCTACAGTCCACAGCACTTTGGGAGCGGAGG	68463		
Qy	1942	CGGAGGATCACTTGAGCCAGGAGTTTGAGACTAGTCTGGGCAACA-----GACACATGT	1997		
Db	68462	CAGAAGGATTCTTGAGCCAGGAGTTCAAGCCAGCGCTGGGCAACATAGCGAGAACCTG	68403		
Qy	1998	CTCTATAGTTGTGTTTGTGTTTGTGTTTACAGGTTGGTGTGTGCACCTGTCAGTCCCA	2057		
Db	68402	TTCTATGAAGATAAATAATTAATAATTAGCCAGCGGTGGTGTGTGCACCTGTGTGCCA	68343		
Qy	2058	GCTACTAGGAGGCTGAGTGGGAGGACTGCTGAGCCAGAGGTGAGGCTGAGTGA	2117		
Db	68342	GCTACTTGGGAGGCTGTGGTAGGAGTGGCTTGAGTCTAGAGTTTGAGGCTGCAGTGA	68283		
Qy	2118	GCCATGATGGCACTGCACTCCAGCCTGGGCAACAGCAAGACCTTGTCTCAAAAC	2177		
Db	68282	GCTATGATGGGCGCGCTGCACTCCAGCCTGGGCAACACAGTGAGACCTGTCTCTAAAT	68223		
Qy	2178	AAACAAAAAGCATACTCATATA	2197		



*This Page Blank (uspto)*

~~*This Page Blank (uspto)*~~

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 22:55:02 ; Search time 411 Seconds  
(without alignments)  
9829.604 Million cell updates/sec

Title: US-10-829-118-2  
Perfect score: 2469  
Sequence: 1 ttgagctctctggaggctc.....gaatacatggcctacagct 2469

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfileseq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	4	US-09-891-711-2
C 2	2452.2	99.3	67156	4	US-09-949-016-12284
C 3	2452.2	99.3	67157	4	US-09-949-016-16558
4	980	39.7	980	4	US-09-891-711-1
5	346	14.0	32010	4	US-09-949-016-13127
C 6	340.4	13.8	40655	4	US-09-949-016-12032
C 7	340.4	13.8	40655	4	US-09-949-016-15919
C 8	332	13.4	187136	4	US-09-949-016-17231
C 9	330.8	13.4	84425	4	US-09-949-016-17402
10	324.2	13.1	111454	4	US-09-949-016-14196
11	324.2	13.1	111459	4	US-09-949-016-12337
C 12	322.6	13.1	43255	4	US-09-949-016-11909
C 13	322.2	13.0	112219	4	US-09-949-016-12453
C 14	322.2	13.0	112222	4	US-09-949-016-14324
C 15	322.2	13.0	113186	4	US-09-949-016-17572
C 16	318	12.9	29067	4	US-09-949-016-12535
C 17	318	12.9	29093	4	US-09-949-016-17476
C 18	317.8	12.9	63760	4	US-09-949-016-14087
C 19	317.8	12.9	63760	4	US-09-949-016-14088
C 20	317	12.8	84587	4	US-09-949-016-15733
C 21	316.8	12.8	89843	4	US-09-949-016-17573
C 22	316.8	12.8	89844	4	US-09-949-016-13656
C 23	316	12.8	63783	4	US-09-949-016-13576
C 24	315.8	12.8	43657	4	US-09-949-016-13777
C 25	314.8	12.8	87562	4	US-09-949-016-13685
C 26	314.2	12.7	24166	4	US-09-949-016-13755
27	314.2	12.7	91665	4	US-09-949-016-12234

ALIGNMENTS

RESULT 1

US-09-891-711-2  
; Sequence 2, Application US/09891711  
; Patent No. 6723553  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Promoter Sequences  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/09/891,711  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2469  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-891-711-2

Query Match 100.0%; Score 2469; DB 4; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TTGAGTCTCTCTGGAGGCTCCGGCCAGAGCAGGCGGTATTGTTTTCACCTCGGTGAATGCT	60
DB	1	TTGAGTCTCTCTGGAGGCTCCGGCCAGAGCAGGCGGTATTGTTTTCACCTCGGTGAATGCT	60
QY	61	CATTTTCACTAAAGAAACACAGGCAACGGAACAAGCTCGGAGCGCGAGACCCCGCA	120
DB	61	CATTTTCACTAAAGAAACACAGGCAACGGAACAAGCTCGGAGCGCGAGACCCCGCA	120
QY	121	GGGCCCGGTACAGGCACTGTGTGTCTCAAAACAGCCCGGAGCCCGCGCCACCTCCC	180
DB	121	GGGCCCGGTACAGGCACTGTGTGTCTCAAAACAGCCCGGAGCCCGCGCCACCTCCC	180
QY	181	CGACCCCGCCCGCCCGCCCGCCCGCTCGCTCGGGGCTCGGACCAACCGGACACCT	240
DB	181	CGACCCCGCCCGCCCGCCCGCCCGCTCGCTCGGGGCTCGGACCAACCGGACACCT	240
QY	241	GAGCAGCGGGCGCCGACACCGCTAGGCGGAGCGGGTCTGGGAGCCCGCGCGGGCTG	300
DB	241	GAGCAGCGGGCGCCGACACCGCTAGGCGGAGCGGGTCTGGGAGCCCGCGCGGGCTG	300
QY	301	ACGTACTCGCCCGCGGAGCTCAGGGCCCGCGGCGCGGATGAGCGGCGCTCCGCG	360
DB	301	ACGTACTCGCCCGCGGAGCTCAGGGCCCGCGGCGCGGATGAGCGGCGCTCCGCG	360
QY	361	CGACACCGCTGGGCACTCCATTCGGGGCTGTTTACTCCCACTCTCGCGAGACTGGCG	420
DB	361	CGACACCGCTGGGCACTCCATTCGGGGCTGTTTACTCCCACTCTCGCGAGACTGGCG	420

Qy		421	GCGGGCCAGCAGGCCCACAGCTGGAGCCTTCAGTCTCGCCGACCAGCGTGCCCTGTGC	480
Db		421	GCGGGCCAGCAGGCCCACAGCTGGAGCCTTCAGTCTCGCCGACCAGCGTGCCCTGTGC	480
Qy		481	TGTTCCCGCGCTCCGGGGCTTGCGTGGCGCCTCTGGACGCCGTGGGCAGCGGGAACAACGC	540
Db		481	TGTTCCCGCGCTCCGGGGCTTGCGTGGCGCCTCTGGACGCCGTGGGCAGCGGGAACAACGC	540
Qy		541	C GGAGGATGGACGAAGGTGCTCGCGACATTTTGGCGGCGCGGGCGCGGTGGCAGGGTGG	600
Db		541	C GGAGGATGGACGAAGGTGCTCGCGACATTTTGGCGGCGCGGGCGCGGTGGCAGGGTGG	600
Qy		601	AAGCGAGGGCGTGGCGACGAGCTGCCAGGCGCGAGAACCGCTGTGGGGGAACCCCTTG	660
Db		601	AAGCGAGGGCGTGGCGACGAGCTGCCAGGCGCGAGAACCGCTGTGGGGGAACCCCTTG	660
Qy		661	GTCCGCTCTGGCGTCCGCTCTAGGATCCCCGAAAGSAGCACGGGCGCGAAGCGGGCAG	720
Db		661	GTCCGCTCTGGCGTCCGCTCTAGGATCCCCGAAAGSAGCACGGGCGCGAAGCGGGCAG	720
Qy		721	GCTGGGCCAGATCTAGAAAGACTGCTTGCGCAGGCTCCCTGCCCCCGGGCCCTGCTG	780
Db		721	GCTGGGCCAGATCTAGAAAGACTGCTTGCGCAGGCTCCCTGCCCCCGGGCCCTGCTG	780
Qy		781	TCATGGACTCGTGGAGAGCTCGCTTCCCGCGCGACACCTTCTGCAAGGCTCCACGTCCA	840
Db		781	TCATGGACTCGTGGAGAGCTCGCTTCCCGCGCGACACCTTCTGCAAGGCTCCACGTCCA	840
Qy		841	GGCACCGCGGCTCGGACACCCACCCCCCGCGCGGACCTGCGCTGGGTGCCCTTTAAC	900
Db		841	GGCACCGCGGCTCGGACACCCACCCCCCGCGCGGACCTGCGCTGGGTGCCCTTTAAC	900
Qy		901	CCGGCGGTAGCTCGTTTAAGATGCGAAGTGTCCGGTCCGGAACACGCGAAAAACCCAAAT	960
Db		901	CCGGCGGTAGCTCGTTTAAGATGCGAAGTGTCCGGTCCGGAACACGCGAAAAACCCAAAT	960
Qy		961	CCCGCTGCCCGACCTCTGACCCCCCGGCCACGGGACGACAGCTGGGCCCTCCGACG	1020
Db		961	CCCGCTGCCCGACCTCTGACCCCCCGGCCACGGGACGACAGCTGGGCCCTCCGACG	1020
Qy		1021	CGCAGCGGCTGCCGGGACACCGGTGGGTGCGAAACGGAGGACCTTTGTAAACGCCAGTG	1080
Db		1021	CGCAGCGGCTGCCGGGACACCGGTGGGTGCGAAACGGAGGACCTTTGTAAACGCCAGTG	1080
Qy		1081	TTTTGCTCTTTTGAABAAAACGAATAAATGTGTAACTGTCTGAAAAGCTTCCGCGCT	1140
Db		1081	TTTTGCTCTTTTGAABAAAACGAATAAATGTGTAACTGTCTGAAAAGCTTCCGCGCT	1140
Qy		1141	AAAAGATGCTCGGTGACTTAGATGCTAGGATCAGTTTGTAACTTTTCAAATGTAATGGACCA	1200
Db		1141	AAAAGATGCTCGGTGACTTAGATGCTAGGATCAGTTTGTAACTTTTCAAATGTAATGGACCA	1200
Qy		1201	CCCGGACTTCGGTACGGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGCTGGTGGGC	1260
Db		1201	CCCGGACTTCGGTACGGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGCTGGTGGGC	1260
Qy		1261	GGTGATCGGCTACAGGCCGATCAGACAGTTTGTGCTTCTGGAACTTGACACTGACCA	1320
Db		1261	GGTGATCGGCTACAGGCCGATCAGACAGTTTGTGCTTCTGGAACTTGACACTGACCA	1320
Qy		1321	CGGTAATGCTGAACTGCACCAATATTACAGATCACAGCGCATCATCTTCTTCAACATGA	1380
Db		1321	CGGTAATGCTGAACTGCACCAATATTACAGATCACAGCGCATCATCTTCTTCAACATGA	1380
Qy		1381	TTTAAACACAGTTGACTTAATATGTGGTGAATAATGTAGAATACAAAATTAACATACCCAC	1440
Db		1381	TTTAAACACAGTTGACTTAATATGTGGTGAATAATGTAGAATACAAAATTAACATACCCAC	1440
Qy		1441	CTCAGGCTTCTACTTCGTAAATTTTGGACAGGTGTTTAACTCTTCTTCTACTCAGCTTCT	1500
Db		1441	CTCAGGCTTCTACTTCGTAAATTTTGGACAGGTGTTTAACTCTTCTTCTACTCAGCTTCT	1500
Qy		1501	TCATTACAAAATAGGGGTACTAGCCAGGGCGGGTGGCTCGCGCTCTGTAATCCCGACNCT	1560

Db	1501	TCATTAACAAAAAGGGGTACTAGCCAGGGGGTGGCTCGCGCTGTAAATCCAGCACT	1560
Qy	1561	TGGGGAGCCGAGCGACGCCGATCACTTGAGGTCAAGAGTTTCAGAGCCAGCTGGTCAAC	1620
Db	1561	TGGGGAGCCGAGCGACGCCGATCACTTGAGGTCAAGAGTTTCAGAGCCAGCTGGTCAAC	1620
Qy	1621	ATGGGTGAACGCCGGCTCTACCAAAATATATAAACTTACGTAGTGTGGTACGCATG	1680
Db	1621	ATGGGTGAACGCCGGCTCTACCAAAATATATAAACTTACGTAGTGTGGTACGCATG	1680
Qy	1681	ACTGTAATCCCAGCACTCAGGAGGCTGAGCGAGAGATCGCTTGAACCTGGAGCGCGGA	1740
Db	1681	ACTGTAATCCCAGCACTCAGGAGGCTGAGCGAGAGATCGCTTGAACCTGGAGCGCGGA	1740
Qy	1741	GGTTGCAGTGAGCTGAGATCGTACCACCTGCACTCCAGCTTGGCGCAGACGAGCACTCT	1800
Db	1741	GGTTGCAGTGAGCTGAGATCGTACCACCTGCACTCCAGCTTGGCGCAGACGAGCACTCT	1800
Qy	1801	GCCTTAAAAATAAATAAATTTTAAAAAATAAGGGTACTTAATCTCCTCTTAAG	1860
Db	1801	GCCTTAAAAATAAATAAATTTTAAAAAATAAGGGTACTTAATCTCCTCTTAAG	1860
Qy	1861	GATGAGGTTTAAATTAAGTACACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTAATCT	1920
Db	1861	GATGAGGTTTAAATTAAGTACACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTAATCT	1920
Qy	1921	CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAGCCCGAGAGTTTGAGACTAGTCT	1980
Db	1921	CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAGCCCGAGAGTTTGAGACTAGTCT	1980
Qy	1981	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTACAGGTGTGGTGGT	2040
Db	1981	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTACAGGTGTGGTGGT	2040
Qy	2041	GTGCACCTGCAGTCCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCCTGAGCCACGGA	2100
Db	2041	GTGCACCTGCAGTCCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCCTGAGCCACGGA	2100
Qy	2101	GGTCGAGGCTGCAGTGAGCCATGATTTGTGCCACTGCACTCCAGCTGGGCAACACAGCAA	2160
Db	2101	GGTCGAGGCTGCAGTGAGCCATGATTTGTGCCACTGCACTCCAGCTGGGCAACACAGCAA	2160
Qy	2161	GACCTTGTCTCAAAAAACAAACAAAAACATACTATAAGTGTCTCGGCTCCTATATGATT	2220
Db	2161	GACCTTGTCTCAAAAAACAAACAAAAACATACTATAAGTGTCTCGGCTCCTATATGATT	2220
Qy	2221	CAATATGTGGTGGATTCCTTGAAATCCTTTTCTGACTCAGATCTCATACGATTTCTGA	2280
Db	2221	CAATATGTGGTGGATTCCTTGAAATCCTTTTCTGACTCAGATCTCATACGATTTCTGA	2280
Qy	2281	ACTTTTGGAGATCCTTTGCCCTCTCTGATTTGCAACCGGTCAAAAGGCACTCCCTCTGCC	2340
Db	2281	ACTTTTGGAGATCCTTTGCCCTCTCTGATTTGCAACCGGTCAAAAGGCACTCCCTCTGCC	2340
Qy	2341	ACCAACAAAGCATTTGATTTTAACTTGACTATGTCTCTTCTGTTCCAACTTTAGGTAAA	2400
Db	2341	ACCAACAAAGCATTTGATTTTAACTTGACTATGTCTCTTCTGTTCCAACTTTAGGTAAA	2400
Qy	2401	TTAATCTTGGTCAGGTTCTCTGAACAGCCCTTTAGTCACCTATGCGCAATTGAATACATGGC	2460
Db	2401	TTAATCTTGGTCAGGTTCTCTGAACAGCCCTTTAGTCACCTATGCGCAATTGAATACATGGC	2460
Qy	2461	CCTACAGCT 2469	
Db	2461	CCTACAGCT 2469	

RESULT 2  
US-09-949-016-12284/c  
; Sequence 12284, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12284
; LENGTH: 67156
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-12284

```

Db 679 GCCTTAAAAATAAATAAATATTTTAAAAAATAGGGGTACTAATAATCTACCTTTAAAG 620  
 Qy 1861 GATGAGGTTAAATTAAGTACACACATAAGCCCTAGCGAGTGGCTTATGCGCTGTAATCT 1920  
 Db 619 GATGAGGTTAAATTAAGTACACACATAAGCCCTAGCGAGTGGCTTATGCGCTGTAATCT 560  
 Qy 1921 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTACAGCCAGGAGTTTGAGACTAGTCT 1980  
 Db 559 CAACACTTTGGGAGTCTGTGGCGGAGGATCACTTACAGCCAGGAGTTTGAGACTAGTCT 500  
 Qy 1981 GGGCAACAGAGCATGCTCTATAGTTGTGTTGGTTTGTGTTTACAGGTTGTGTTGT 2040  
 Db 499 GGGCAACAGAGCATGCTCTATAGTTGTGTTGGTTTGTGTTTACAGGTTGTGTTGT 440  
 Qy 2041 GTGCACCTGAGTCCAGCTACTAGGAGGCTGAGTGGGAGACTGCTGAGCCAGCA 2100  
 Db 439 GTGCACCTGAGTCCAGCTACTAGGAGGCTGAGTGGGAGACTGCTGAGCCAGCA 380  
 Qy 2101 GGTGAGGCTGACGTAGGAGGATGATGTCACCTGACCTCCAGCCTGGGCAACACAGCAA 2160  
 Db 379 GGTGAGGCTGACGTAGGAGGATGATGTCACCTGACCTCCAGCCTGGGCAACACAGCAA 320  
 Qy 2161 GACCTTGTCTCAAAAACAAAAGCATACTATAAAGTGTCTCGGCTCCTATATGATT 2220  
 Db 319 GACCTTGTCTCAAAAACAAAAGCATACTATAAAGTGTCTCGGCTCCTATATGATT 260  
 Qy 2221 CAATATGCTGTGATGATCTTGAATCTTCTGACTCAGATCTCATAGGATTTCTGA 2280  
 Db 259 CAATATGCTGTGATGATCTTGAATCTTCTGACTCAGATCTCATAGGATTTCTGA 200  
 Qy 2281 ACTTTTGGAGATCTTGTGCTCTCTGATTTGCAACCGTCAAGGACACTCCCTTCTGCC 2340  
 Db 199 ACTTTTGGAGATCTTGTGCTCTCTGATTTGCAACCGTCAAGGACACTCCCTTCTGCC 140  
 Qy 2341 ACCACAAAGCATTTGATTTTAACTTGAATATGATGTCCTTCTGTTCCAACTTTAGGTAAA 2400  
 Db 139 ACCACAAAGCATTTGATTTTAACTTGAATATGATGTCCTTCTGTTCCAACTTTAGGTAAA 80  
 Qy 2401 TTAATCTTGTGAGGTTCTGAAACAGCCCTTATGATGATGATGATGATGATGATGATGATG 2460  
 Db 79 TTAATCTTGTGAGGTTCTGAAACAGCCCTTTAGTCACTATGCAATGCAATGCAATGCA 20  
 Qy 2461 CCTACAGCT 2469  
 Db 19 CCTACAGCT 11

RESULT 3  
 US-09-949-016-16558/c  
 / Sequence 16558, Application US/09949016  
 / Patent No. 6812339  
 / GENERAL INFORMATION:  
 / APPLICANT: VENTER, J. Craig et al.  
 / TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 / WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 / FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 / FILE REFERENCE: CL001307  
 / CURRENT APPLICATION NUMBER: US/09/949, 016  
 / PRIOR FILING DATE: 2000-04-14  
 / PRIOR APPLICATION NUMBER: 60/241,755  
 / PRIOR FILING DATE: 2000-10-20  
 / PRIOR APPLICATION NUMBER: 60/237,768  
 / PRIOR FILING DATE: 2000-10-03  
 / PRIOR APPLICATION NUMBER: 60/231,498  
 / PRIOR FILING DATE: 2000-09-08  
 / NUMBER OF SEQ ID NOS: 207012  
 / SOFTWARE: FastSeq for Windows Version 4.0  
 / SEQ ID NO 16558  
 / LENGTH: 67157  
 / TYPE: DNA  
 / ORGANISM: Human  
 US-09-949-016-16558  
 Query Match 99.3%; Score 2452.2; DB 4; Length 67157;

Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 2465; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
 Qy 1 TTGAGTCTGTCTGGAGGCTCCGGGCGAGAGCGGGGTATTTCTTTCACCTCGGTGATGCT 60  
 Db 2478 TTGAGTCTGTCTGGAGGCTCCGGGCGAGAGCGGGGTATTTCTTTCACCTCGGTGATGCT 2419  
 Qy 61 CATTTTCAGTAAAGAAAACACGAGCAACAAAGCTGCGGAGCGCGAGACCCCGCA 120  
 Db 2418 CATTTTCAGTAAAGAAAACACGAGCAACAAAGCTGCGGAGCGCGAGACCCCGCA 2359  
 Qy 121 GGGCCGCGGTACAGGCAAGCTGTGTCTCAAAAGCGCGGAGGCGCCCGCCACCTCC 180  
 Db 2358 GGGCCGCGGTACAGGCAAGCTGTGTCTCAAAAGCGCGGAGGCGCCCGCCACCTCC 2299  
 Qy 181 CGGACCGGCG 240  
 Db 2298 CGGACCGGCG 2239  
 Qy 241 GAGCGAGCGGCG 300  
 Db 2238 GAGCGAGCGGCG 2179  
 Qy 301 AGGTACTGCG 360  
 Db 2178 AGGTACTGCG 2119  
 Qy 361 CGAACAAGCG 420  
 Db 2118 CGAACAAGCG 2059  
 Qy 421 GCGGCG 480  
 Db 2058 GCGGCG 1999  
 Qy 481 TGTCTCCG 540  
 Db 1998 TGTCTCCG 1939  
 Qy 541 CGGAGGATGGAAGAGTGTCTCGGAGCAATTTGCGGCGCGCGCGCGCGCGCGCGCG 600  
 Db 1938 CGGAGGATGGAAGAGTGTCTCGGAGCAATTTGCGGCGCGCGCGCGCGCGCGCGCG 1880  
 Qy 601 AAGCGGAGGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
 Db 1879 AAGCGGAGGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1820  
 Qy 661 GTCCGCTCTGCGCGGTCTTAGGATCCCGGAAAAGGAGCAGCGGCGCGGAAAGCGGCCAG 720  
 Db 1819 GTCCGCTCTGCGCGGTCTTAGGATCCCGGAAAAGGAGCAGCGGCGCGGAAAGCGGCCAG 1760  
 Qy 721 GTTGGGCGAGGATCTAGAAAAGCTGCTGCGGAGAGGCTCCCTGCGCGCGCGCGCGCG 780  
 Db 1759 GCTTGGGCGAGGATCTAGAAAAGCTGCTGCGGAGAGGCTCCCTGCGCGCGCGCGCGCG 1700  
 Qy 781 TCATGAGCTCTGAGAGCTCTCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840  
 Db 1699 TCATGAGCTCTGAGAGCTCTCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1640  
 Qy 841 GGCACCGCGCGGTCTGAGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
 Db 1639 GGCACCGCGCGGTCTGAGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1580  
 Qy 901 CGGCGCGGTAGCTGTGTAAGATGCGGAGTGTCTCGGTCGCGGACACCGCGGAAACCCCAAT 960  
 Db 1579 CGGCGCGGTAGCTGTGTAAGATGCGGAGTGTCTCGGTCGCGGACACCGCGGAAACCCCAAT 1520  
 Qy 961 CGGCGCGGTAGCTGTGTAAGATGCGGAGTGTCTCGGTCGCGGACACCGCGGAGCTGCGGCG 1020  
 Db 1519 CGGCGCGGTAGCTGTGTAAGATGCGGAGTGTCTCGGTCGCGGACACCGCGGAGCTGCGGCG 1460  
 Qy 1021 CGGCGCGGTAGCTGTGTAAGATGCGGAGTGTCTCGGTCGCGGACACCGGAGGACCTTTGTAACGCGCCAGCTG 1080

1459 CGCAGCGCGTCTGCGGAGACACCGGTGCTGCGAAGCGAGGACCTTTGTAAAGCCACGCTG 1400  
1081 TTTGCTCTTTTGAAGAAACAGAAATAATGCTTAACTGCTGAAGAGCTTGGCGCCT 1140  
1399 TTTGCTCTTTTGAAGAAACAGAAATAATGCTTAACTGCTGAAGAGCTTGGCGCCT 1340  
1141 AAAAGATGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1200  
1339 AAAAGATGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1280  
1201 CCCGGAATGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1260  
1279 CCCGGAATGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1220  
1261 GGTGATGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1320  
1219 GGTGATGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1160  
1321 CGGTAATGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1380  
1159 CGGTAATGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1100  
1381 TTTAACACAGTTCGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1440  
1099 TTTAACACAGTTCGCTGCGGTGACCTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGAACAG 1040  
1441 CTGAGGCTTCTACTTCTGTAATTTTGAAGAGTGTGTTTAACTCTTGTACCTCAGCTTCT 1500  
1039 CTGAGGCTTCTACTTCTGTAATTTTGAAGAGTGTGTTTAACTCTTGTACCTCAGCTTCT 980  
1501 TCATTACAAAATAGGGGTACTAGCAGGCGGGTGGCTCGCGCTGTATCCAGCACT 1560  
979 TCATTACAAAATAGGGGTACTAGCAGGCGGGTGGCTCGCGCTGTATCCAGCACT 920  
1561 TGGGAGGCGCGAGCGAGCGGATCCTTGAAGTGAAGTTTTCAGACCGCTGGTCAAC 1620  
919 TGGGAGGCGCGAGCGAGCGGATCCTTGAAGTGAAGTTTTCAGACCGCTGGTCAAC 860  
1621 ATGGGTGAACCGCGGCTCTACCAAAATATAAATTAAGTGAAGTTTTCAGACCGCTGGTCAAC 1680  
859 ATGGGTGAACCGCGGCTCTACCAAAATATAAATTAAGTGAAGTTTTCAGACCGCTGGTCAAC 800  
1681 ACTGTAATCCAGCACTCAGGAGGCTGAGCGAGAGTTCGTTGAACCTGGGAGCGGA 1740  
799 ACTGTAATCCAGCACTCAGGAGGCTGAGCGAGAGTTCGTTGAACCTGGGAGCGGA 740  
1741 GGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1800  
739 GGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 680  
1801 GCCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAG 1860  
679 GCCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAG 620  
1861 GATGAGGTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 1920  
619 GATGAGGTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 560  
1921 CAAACATTTGGGAGTCTGTGCGGAGGATCCTTGAAGCCAGAGTTTGAAGTGAAGTGAAGTGAAG 1980  
559 CAAACATTTGGGAGTCTGTGCGGAGGATCCTTGAAGCCAGAGTTTGAAGTGAAGTGAAGTGAAG 500  
1981 GGGCAACAGAGACATGCTCTATAGTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 2040  
499 GGGCAACAGAGACATGCTCTATAGTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 440  
2041 GTGACCTGCGAGTCCAGCTACTAGGAGGCTGAGTGGAGGATGCTGCTGAGCCAGGA 2100  
439 GTGACCTGCGAGTCCAGCTACTAGGAGGCTGAGTGGAGGATGCTGCTGAGCCAGGA 380  
2101 GGTGAGGCTGCGAGTCCAGCTACTAGGAGGCTGAGTGGAGGATGCTGCTGAGCCAGGA 2160  
379 GGTGAGGCTGCGAGTCCAGCTACTAGGAGGCTGAGTGGAGGATGCTGCTGAGCCAGGA 320

2161 GACCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2220  
319 GACCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 260  
2221 CAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280  
259 CAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 200  
2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAAAACCTGCAAAAGCACTCCCTTCTGCC 2340  
199 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAAAACCTGCAAAAGCACTCCCTTCTGCC 140  
2341 ACCACACAAAGCAATTTGATTTTAACTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400  
139 ACCACACAAAGCAATTTGATTTTAACTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 80  
2401 TTAATCTTGGTCAAGGTTCTCTGCAAGAGGCTTTAGTCACTATGCAATGCAATGCAATGCA 2460  
79 TTAATCTTGGTCAAGGTTCTCTGCAAGAGGCTTTAGTCACTATGCAATGCAATGCAATGCA 20  
2461 CCTACAGCT 2469  
19 CCTACAGCT 11

RESULT 4  
US-09-891-711-1  
; Sequence 1, Application US/09891711  
; Patent No. 6723553  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: promoter Sequences  
; FILE REFERENCE: 00130  
; CURRENT APPLICATION NUMBER: US/09/891,711  
; CURRENT FILING DATE: 2001-06-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 980  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-891-711-1

Query Match 39.7%; Score 980; DB 4; Length 980;  
Best Local Similarity 100.0%; Pred. No. 7.2e-230;  
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

353 CTCGCGCGGAAACACGCTGGGCACTCCATTTCGGGCTGTTTACTCCCAACTCTCGCGAG 412  
1 CTCGCGCGGAAACACGCTGGGCACTCCATTTCGGGCTGTTTACTCCCAACTCTCGCGAG 60  
413 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 472  
61 ACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
473 GCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 532  
121 GCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
533 GACCAACCGCGGAGGATGGAACGAGTGTCTCGGACATTTTGGCGGCGGCGGCGGCGGCGGCGG 592  
181 GACCAACCGCGGAGGATGGAACGAGTGTCTCGGACATTTTGGCGGCGGCGGCGGCGGCGGCGG 240  
593 CAGGCTGGAAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 652  
241 CAGGCTGGAAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
653 AACCTTGGTCCGCTCTGCGGCTGCTGAGGATCCCGAAAGGAGCAGCGGCGGCGGCGGCGG 712  
301 AACCTTGGTCCGCTCTGCGGCTGCTGAGGATCCCGAAAGGAGCAGCGGCGGCGGCGGCGG 360  
713 GCGGCCAGGCTGGGCGGAGGATCTAGAAAGACTGCTGGCGGAGGCTCCCTGCCCCCGCGG 772



```

Db 361 GCGGCGAGGCTGGCCAGGATCTAGAAAGATCTGCTGGCGCAGGCTCCCTGCCCCCGCG 420
QY 773 GCTGTGTATGATGCTGCGGAGAGCTGCTTTCCGCGCGGACCTTCTCTGAGGGGTC 832
Db 421 GCTGTGTATGATGCTGCGGAGAGCTGCTTTCCGCGCGGACCTTCTCTGAGGGGTC 480
QY 833 CAGTCCAGGACCGGGCTCGGACACCCACCCCGCGCGGACCTGCGCTGGGTGC 892
Db 481 CAGTCCAGGACCGGGCTCGGACACCCACCCCGCGCGGACCTGCGCTGGGTGC 540
QY 893 CCCTTAACCGGGCTGAGCTGCTTAAGATGGGAAAGTCTCCGGTCCGGAAACACGCGAAA 952
Db 541 CCCTTAACCGGGCTGAGCTGCTTAAGATGGGAAAGTCTCCGGTCCGGAAACACGCGAAA 600
QY 953 CCCAAATCCCGCTGCGGACCTCTGACCCCGCGCGGACGAGACGAGACTGGGCC 1012
Db 601 CCCAAATCCCGCTGCGGACCTCTGACCCCGCGCGGACGAGACGAGACTGGGCC 660
QY 1013 TCCGAGCGGACGGGCTGCGGAGACCGGTGCGTGGGAAACGAGGACCTTTGTAAC 1072
Db 661 TCCGAGCGGACGGGCTGCGGAGACCGGTGCGTGGGAAACGAGGACCTTTGTAAC 720
QY 1073 GCCACGTGTTGCTCTTTTGAAGAAACAAAGATAAATGTTTAAACTGTCTGAAAAGCT 1132
Db 721 GCCACGTGTTGCTCTTTTGAAGAAACAAAGATAAATGTTTAAACTGTCTGAAAAGCT 780
QY 1133 TGCGGCTAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGAAA 1192
Db 781 TGCGGCTAAAGATGCTGGGTGACTTAGATGCTAGGATCAGTTGTTTCAATGAAA 840
QY 1193 TGACACCGCGGACCTCCGTACGGCACTAGCAGGAGCTGAAAGCTCTTCAGTACTGC 1252
Db 841 TGACACCGCGGACCTCCGTACGGCACTAGCAGGAGCTGAAAGCTCTTCAGTACTGC 900
QY 1253 TGGTGGCGGTGATGCGCTACAGCGGATCAGACAGTTTGTGCTCTTCTGAACTTGACA 1312
Db 901 TGGTGGCGGTGATGCGCTACAGCGGATCAGACAGTTTGTGCTCTTCTGAACTTGACA 960
QY 1313 CTGCACACGGTAAATGCTGA 1332
Db 961 CTGCACACGGTAAATGCTGA 980

```

RESULT 5

```

US-09-949-016-13127
; Sequence 13127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13127
; LENGTH: 32010
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13127

```

Query Match 14.0%; Score 346; DB 4; Length 32010;  
 Best Local Similarity 71.4%; Pred. No. 3.3e-74;  
 Matches 510; Conservative 0; Mismatches 200; Indels 4; Gaps 4;

```

QY 1480 CCTCTTTGTACTCAGCTTCTTTTACAAAAATAGGGTACTAGCCAGCGGGGTGGCT 1539
Db 6160 CTTCTTTGTAGCTCATCTGCAAAAATGTGGATCAATCACCTTTGGGCGAGGCCGGTGGCT 6219
QY 1540 CCGCGCTGTAAATCCAGCACTTTGGGAGGCCAGGCGGATCACTTTGAGGTCAAGAAG 1599
Db 6220 CACACCTGTAAATCCAGCGCTTTGGAGGCCAAGCGAGGTAGATCACTTTGAGGCCAGGAG 6279
QY 1600 TTTACAGCAGCGCTGGTCAACATGGGTGAACCGCCGGCTCTACCAAAAAATATAAAAACTT 1659
Db 6280 TTTGGAGCAGCGCTGGCCAAACAT-GGTGAAACCCCGTCTCTACCAAAAAACAAAAAATTT 6338
QY 1660 AGCTGAGTGTGTAGCGCATGACTGTAATCCAGCAACTCAGAGGCTCAGGCAG-AGAA 1718
Db 6339 AGCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6398
QY 1719 TCGCTTGAACCTGGGAGCGGAGGTTGCGAGTGTGAGTGTGAGATCGTACCTGCACTCCAGC 1778
Db 6399 TTTGCTTAAACCTGGGAGCGGAGGTTGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 6458
QY 1779 TTTGGGCGCAGAGCGAGACTCTGCCTTAAAAATTAATAATAATAATAATAATAATAATAAG 1838
Db 6459 CTGGGCAACAGAGCGAGACTCTGTCTAAAAAATAAATAAATAAATAAATAAATAAATAA 6518
QY 1839 GGTACTAATATCTACCTTAAAGGATGAGGTTTAAATTAAGTACACACATAAAGCCCTAGCG 1898
Db 6519 CCATGCTCTCTATGCCATGGTTGCTGTGAGGACTTGGAAATAGCAC-CTGGTGGCGGGCG 6577
QY 1899 CAGTGGCTTATGCTGTATCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTTGAG 1958
Db 6578 CAGTGTCTCACGCTGTATCTCCAGCACCTTTGGGAGGCTGAGCGGCGAGATCACTTTGAG 6637
QY 1959 CCGAGAGTTTGAAGTAGTAGTCTGGGCAACAGAGACATGTCTCTATAGTTGTGTGGTTT 2018
Db 6638 GCCAGAAGTTTGAAGTAGTAGTCTGGGCAACAGAGACATGTCTCTCTCTCTCTCTCTCT 6697
QY 2019 TGTCTTTACAGGTGTGGTGTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGGTG 2078
Db 6698 AAAATTAGCTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 6757
QY 2079 GGAGGACTGCTGAGCCAGGAGGTTGAGGCTGCGAGTGTGAGTGTGAGTGTGAGTGTGAG 2138
Db 6758 GGAGAATCACTTTGAATCTCGGAGGCGGAGGTTTGCAGTGTGAGTGTGAGTGTGAGTGTG 6817
QY 2139 TCGAGCTGGGCAAC-ACAGCAAGACCTTGTCTCAAAAAACAAAAAGCATA 2191
Db 6818 TCCAGCTGGGCGCAACAGCAAGACCTTGTCTCTCAAAAAACAAAAAGCATA 6871

```

RESULT 6

```

US-09-949-016-12032/c
; Sequence 12032, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12032
; LENGTH: 40655
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12032

```



```

Query Match      13.8%; Score 340.4; DB 4; Length 40655;
Best Local Similarity 71.8%; Pred. No. 8.7e-73;
Matches 501; Conservative 0; Mismatches 191; Indels 6; Gaps 4;

QY 1508 AAAAAATAGGGTACTAGCAGCGGGGTGGCTCGCGCTGTAAATCCAGCACTTGGGGAG 1567
DB 10739 AGAATTAAGTATATTGGCCAGGCGATGGTGGCTCAGCCCAATAATCACAGGACTTTGGGAG 10680
QY 1568 GCGAGGAGCGGATCACTTGAAGTTCAGAACCTTTCAGACCGGCTGGTCAACATGGGTG 1627
DB 10679 GCGAAGGAGGCGAGATCACTGAGGTTCAGAGGTTTGAGACTAGCCTGACAAACAT-GGTA 10621
QY 1628 AAACGCGGCTCTACCAAAAAT--ATAAAAACTTAGCTGAGTGTGGTGGCATCACTGT 1685
DB 10620 AAACCTGTCTTACTTAAATAACAAAAAATAGCCGGTGTGGTGGCAGACACCTGT 10561
QY 1686 AATCCAGCAACTCAGAGGCTGAGCA-GAGAAATCGTTGAACCTTGGAGGCGGAGGTT 1744
DB 10560 AATCCAGCTACTCAGAGGCTGAGGAGGAGAAATCGTTGAACCGAGAGGTGGAGGTT 10501
QY 1745 GCAGTGAAGTGAATCGTACCACTGCACTCCAGCTTGGGCGAGAGGAGGACTCTGCT 1804
DB 10500 GCAGTGAAGTGAATCGTACCACTGCACTCCAGCTTGGGCGAGAGGAGGACTCTGCT 10441
QY 1805 TAAAAATAAATAATATTTTAAAAAATAAGGGTACTTAATATCTACCTTAAAGG--A 1862
DB 10440 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 10381
QY 1863 TGAGGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCA 1922
DB 10380 TATATGTTAAATTAAGTACACATAGCCCTAGCGAGGAGGAGGAGGAGGAGGAGG 10321
QY 1923 AACTTTGGAGGCTGTGGCGGAGGAGTCACTAGCGAGGAGGAGGAGGAGGAGGAGGAGG 1982
DB 10320 GCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10261
QY 2043 GCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2102
DB 10200 GTGCTATAATCTCAGCTACTAGGAGGCTGAGTGGAGGAGGAGGAGGAGGAGGAGG 10141
QY 2103 TCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2162
DB 10140 TCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10081
QY 2163 CCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 2200
DB 10080 CACTGCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 10043

```

```

RESULT 7
US-09-949-016-15919/c
; Sequence 15919, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 15919
; LENGTH: 40655
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-15919

Query Match      13.8%; Score 340.4; DB 4; Length 40655;
Best Local Similarity 71.8%; Pred. No. 8.7e-73;
Matches 501; Conservative 0; Mismatches 191; Indels 6; Gaps 4;

QY 1508 AAAAAATAGGGTACTAGCAGCGGGGTGGCTCGCGCTGTAAATCCAGCACTTGGGGAG 1567
DB 10739 AGAATTAAGTATATTGGCCAGGCGATGGTGGCTCAGCCCAATAATCACAGGACTTTGGGAG 10680
QY 1568 GCGAGGAGCGGATCACTTGAAGTTCAGAACCTTTCAGACCGGCTGGTCAACATGGGTG 1627
DB 10679 GCGAAGGAGGCGAGATCACTGAGGTTCAGAGGTTTGAGACTAGCCTGACAAACAT-GGTA 10621
QY 1628 AAACGCGGCTCTACCAAAAAT--ATAAAAACTTAGCTGAGTGTGGTGGCATCACTGT 1685
DB 10620 AAACCTGTCTTACTTAAATAACAAAAAATAGCCGGTGTGGTGGCAGACACCTGT 10561
QY 1686 AATCCAGCAACTCAGAGGCTGAGCA-GAGAAATCGTTGAACCTTGGAGGCGGAGGTT 1744
DB 10560 AATCCAGCTACTCAGAGGCTGAGGAGGAGAAATCGTTGAACCGAGAGGTGGAGGTT 10501
QY 1745 GCAGTGAAGTGAATCGTACCACTGCACTCCAGCTTGGGCGAGAGGAGGACTCTGCT 1804
DB 10500 GCAGTGAAGTGAATCGTACCACTGCACTCCAGCTTGGGCGAGAGGAGGACTCTGCT 10441
QY 1805 TAAAAATAAATAATATTTTAAAAAATAAGGGTACTTAATATCTACCTTAAAGG--A 1862
DB 10440 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 10381
QY 1863 TGAGGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCA 1922
DB 10380 TATATGTTAAATTAAGTACACATAGCCCTAGCGAGGAGGAGGAGGAGGAGGAGGAGG 10321
QY 1923 AACTTTGGAGGCTGTGGCGGAGGAGTCACTAGCGAGGAGGAGGAGGAGGAGGAGGAGG 1982
DB 10320 GCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10261
QY 1983 GCAACAGAGACATGCTCTATAGTGTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2042
DB 10260 GCAACATGCGGAAACCCCATCTCTACCAAAAATAACAAAAATTAGCCATGATGATG 10201
QY 2043 GCACCTGAGTCCAGCTACTAGGAGGCTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGG 2102
DB 10200 GTGCTATAATCTCAGCTACTAGGAGGCTGAGTGGAGGAGGAGGAGGAGGAGGAGG 10141
QY 2103 TCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2162
DB 10140 TCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10081
QY 2163 CCTTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 2200
DB 10080 CACTGCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 10043

```

```

RESULT 8
US-09-949-016-17231/c
; Sequence 17231, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```

; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 17231  
 ; LENGTH: 187136  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(187136)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ;  
 US-09-949-016-17231

Query Match 13.4%; Score 332; DB 4; Length 187136;  
 Best Local Similarity 70.8%; Pred. No. 1.9e-70;  
 Matches 487; Conservative 0; Mismatches 190; Indels 11; Gaps 3;

QY	1508	AAAAATAGGGGTACTAGCCAGGCGGGGTGGCTCGCGCCTGTATCCAGCACTTGGGGAG	1567
DB	24550	AAATATTCAGAAAGAGCCAGGCGGGTGTTCATGCTGTAAACCCAGCACTTGGGGAG	24491
QY	1568	GCCGAGGCGAGCGGATCACTTGAGGTGAGAAATTCAGACCAGCCTGGTCAACATGGGTG	1627
DB	24490	GCTGAGGTGGCGGATGACCTAAGGTGAGCAGTTTGAGACCAGCCTGGCCACAT--GGTG	24432
QY	1628	AAACGCGCGCTTACCAAAATATAAAAACTTAGCTAGGTGTGGTAGCGCATGCTGTAA	1687
DB	24431	AAACCTGTCTTACTTAAAAATACAAAAGTTAGCTGAGCATGGTGGCTCACATCTGTAA	24372
QY	1688	TCCAGCACTCAGGAGGCTGAGGCA--GAGAACTCGCTTGAACCTGGAGCGCGAGTTGC	1746
DB	24371	TCCAGCTACTCGGAGGGGTGAGCGAGGAGAAATCGCTTGAACCCAGGAAGCAGAGTTGC	24312
QY	1747	AGTGAGCTGAGATCGTACCACCTGCATCCAGCTTTGGCGCAGCAGCGAGCTCTGCCTTA	1806
DB	24311	AGTGAGCAGAGATCGCACCATTCATCTCAGCTTGGCGCAACAGCAAGCATCTCGTCTCA	24252
QY	1807	AAAAATAAATAAATTTTTTAAAAAATAAGGGTACTAATATCTACCTTAAAGGATGAG	1866
DB	24251	AAAAAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAATATAAAA	24192
QY	1867	GGTTAAATTA-----AGTACACATAAGCCCTAGCGCAGTGGCTTATGCCGTAA	1917
DB	24191	GCAAGCATTTAAAAAATAAATAAATAAATAAAGGCCAGGCACAGTGGCTCATAGCTGTAA	24132
QY	1918	TCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTGAGCCAGAGTTTGAGACTAG	1977
DB	24131	TCCAGCACTTCGGGAGGCGAGCGAGCGAGATCACTGAGGTGAGGGTTGAGACCAG	24072
QY	1978	TCTGGGCAACAGAGACATGTCTATAGTTGTGTTTGGTTTGTGTTTACCAGGTGGT	2037
DB	24071	CCTGGCCAACTGGTGAACCTGTCTCTACAAAAAATAAAAAAATAGCCAGGCGTGT	24012
QY	2038	GGTGTGCACTGAGTCCAGCTACTAGGAGGCTGAGGTGGAGAGTGCCTGAGCCCA	2097
DB	24011	GGTGCATGCTTAAATCCCAAGCTACTTGGGAGACTGAGGAGAGGAATCACTTGAACCCG	23952
QY	2098	GGAGGTGCGAGGCTGAGTGGAGCATGATTTGTCACCTGCACTCCAGCTGGGCAACACAG	2157
DB	23951	AGAGATGAGAGTTGCGATGAGCTGAGATTGCGGCATGCACTCCAGCTGGCCCAACAG	23892
QY	2158	CAAGACTTTGTCTCAAAAAACAAACAAA	2185
DB	23891	TGAGACTTTGTCTCAAAAAAATAAAAAA	23864

RESULT 9

US-09-949-016-17402/c  
 ; Sequence 17402, Application US/09949016  
 ; Patent No. 681239  
 ; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 17402  
 ; LENGTH: 84425  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ;  
 US-09-949-016-17402

Query Match 13.4%; Score 330.8; DB 4; Length 84425;  
 Best Local Similarity 72.3%; Pred. No. 2.6e-70;  
 Matches 506; Conservative 0; Mismatches 172; Indels 22; Gaps 5;

QY	1494	AGCTTCTTCAATACAAAAATAGGGGTACTAGCCAGGCGGGTGGCTCGCGCCTGTAAATCC	1553
DB	39272	AGATTCTGTCTTAAATAAAAAAGATAACGGCGCGCGGTGTTTCATGCTGCTGTAAATCC	39213
QY	1554	CAGCACTTGGGAGGCGGAGGCGGATCACTTTGAGGTGAGAAATTTAGAGCCAGCCT	1613
DB	39212	CAGCACTTGGGAGGCGGAGGCGGATCAC--GAGGTGAGGAGATTTGAGACCATCTCT	39155
QY	1614	GGTCAACATGGGTGAAACCGCGCTCTACCAAAATAATAAAAACTTAGCTGAGTGGTA	1673
DB	39154	GGTTAAACAT--GGTGAACCCCGTCTCTACTAAAAATACAAAAATTTAGCTGGCGGTG	39096
QY	1674	GGCATGACTGTAAATCCAGCAACTCAGGAGGCTGAGGCA--GAGAACTGCTTGAACCTGG	1732
DB	39095	GCAGTGTCTTGTAGTCTCAGCTACTCAGAGGCTGAGGAGGAGATGGGTGNAACCCG	39036
QY	1733	GAGGCGGAGTTTGCAGTGTAGTCTGATCGTACCTGCACTCCAGCTTGGGCGACAGAGC	1792
DB	39035	GAGGCGGAGTTTGCAGTGTAGTCTGATCGTGCCTGCACTCCAGCTGGGCGACAGAGC	38976
QY	1793	GAGACTCTGCTTAAAAATAAATAAATAAATTTTAAAAAATAAGGGTACTAATATCTA	1852
DB	38975	AAGACTCCGCTCTCAAAAAAATAAAAAATAAATAAATAA-----TAA	38932
QY	1853	CCTTAAAGGATGAGGGTTAAATTAAGTACACACATAAGCCCTAGCGCAGTGGCTTATGCC	1912
DB	38931	ATAAAAAAGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	38872
QY	1913	TGTAATCTCAACACTTTGGGAGTCTGTGGCGGAGGATCACTTTGAGCCCAAGGATTTGAG	1972
DB	38871	TGTAATCCAGCAACTTTGGGAGGCTGATGGGCGCGGATCACTTTAAGGTGAGGATTTGAG	38812
QY	1973	ACTAGTCTGGGCAACA--GAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTACCAG	2030
DB	38811	ACCAGCTGGCCAACTATGTTGTAACCTATCTCAATAAATAAATAAATAAATAAATAA	38752
QY	2031	GTGTGGTGTGTGCACCTGCAGTCCAGCTACTTAGGGAGGCTGAGGTGGGAGGACTGCCT	2090
DB	38751	ATGTGTGTGTGTGCACCTGTATCCAGATACTCAGAGGCTGAGGAGCAGCAAGATCGCTC	38692
QY	2091	GAGCCCAAGGAGTTCGAGGCTGAGGATGAGCCATGATTTGTGCCACTGCACCTCCAGCTGGC	2150
DB	38691	GTACCCAGGAGGCGAGAGTTGCAGTGAGCCCAAGATCTAGCCACTGCACCTCCAGCTGGAT	38632
QY	2151	AACACAGCAAGACCTTGTCTCAAAAAACAAACAAAAGCAT	2190
DB	38631	GACAAAGTGAGACTTTTGTCTCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAA	38592

RESULT 10					
US-09-949-016-14196					
; Sequence 14196, Application US/09949016					
; Patent No. 6812339					
; GENERAL INFORMATION:					
; APPLICANT: VENTER, J. Craig et al.					
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED					
; FILE REFERENCE: CL001307					
; CURRENT FILING DATE: 2000-04-14					
; PRIOR APPLICATION NUMBER: 60/241,755					
; PRIOR FILLING DATE: 2000-10-20					
; PRIOR APPLICATION NUMBER: 60/237,768					
; PRIOR FILLING DATE: 2000-10-03					
; PRIOR APPLICATION NUMBER: 60/231,498					
; PRIOR FILLING DATE: 2000-09-08					
; NUMBER OF SEQ ID NOS: 207012					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 14196					
; LENGTH: 111454					
; TYPE: DNA					
; ORGANISM: Human					
; FEATURE:					
; NAME/KEY: misc_feature					
; LOCATION: (1)...(111454)					
; OTHER INFORMATION: n = A,T,C or G					
US-09-949-016-14196					
Query Match 13.1%; Score 324.2; DB 4; Length 111454;					
Best Local Similarity 70.5%; Pred. No. 1.2e-68;					
Matches 514; Conservative 0; Mismatches 188; Indels 27; Gaps 5;					
Qy	1509	AAATAGGGGTACTAGCCAGCGGGTGGCTCGGCTGTGAATCCACAGCACTTTGGGGAGG	1568		
Dd	37650	AAAAATGGAGACTTGGCGGGCGAGTGCGCTGTGAATCGCAGCACATTGGGAGA	37709		
Qy	1569	CCGAGGCAGCGCATCTTAGGTTCAGAGTTTTCAGACCAGCCTGGTCAACATGGGTGA	1628		
Dd	37710	CTGAGCGGGTGGATCATTGAGGTTCAGGATTCAGAGCCAGCCTAGCCAACAT-GGTGA	37768		
Qy	1629	AACGCCGGCTCTACCAAAAATAAATAAACATTAGCTGAGTGTGTAGGCAATGCTTAAT	1688		
Dd	37769	AACTTTGTCTACTAAATAATACAATAAATTAGCTGGTGTGGTCATCGCCTGTAAT	37828		
Qy	1689	CCCAGCAACTCAGGAGCTGAGCA-----CAGATCGCTTGAACCTGGGAGCGGAGTT	1744		
Dd	37829	CCAGTTTACTCAGGAGCTCAGGAGGAGAGAAATGCTTGAACCCAGAGGTGGAGTT	37889		
Qy	1745	GCACTGAGCTGAGATCGTACCACTCCAGCTTGGGCGACAGACAGCACTTGCCT	1804		
Dd	37889	TCACTGAGCTGAGATCATGCACTATCTACTCCAGCTTGGGCGACAGAGCGAGATTGCCT	37948		
Qy	1805	TAAAAATAAATAATAATTTTTTAAAAAATAAGGGTACTTAATATCTACTTAAAGGATG	1864		
Dd	37949	AAAAAACAA--AAACAACAATAAATAAGAGGGGTGGGACTTAATCATAAAGAAATA	38006		
Qy	1865	AGGTTAAATTAAGTACACATAAGCCCTAGCCAGTGCCTTATGCTGTATCTCAAC	1924		
Dd	38007	GTTATTTGC-----AGGCCAGGCGCAGTGCTTATACCTGTAAATCCAGT	38051		
Qy	1925	ACTTTGGAGTCTGTGGCGGAGGATCATTGAGCCCCAGAGTTTGAGACTACTCTGGGC	1984		
Dd	38052	GCTTTTGGAGGCCAAGTTGGAGAGATTCTTGTAGCCCCAGAGTTCAAGACCGCTTGGC	38111		
Qy	1985	AACAGAGACA-----TGCTCTCTATAGTTGTGTTTGGTTTGTTTTACCAGGTGTGGTG	2039		
Dd	38112	AACATAGCAAGACCCTGTCTACAAAAGTAAAAAATAAAAAAAAAAAGTCAAGCGGTAG	38171		
Qy	2040	TGTGCACTTGAGTCCAGCTTACTAGGAGGCTGAGGTGGAGGACTGCTAGGCCAGG	2099		
Dd	38172	TGTGGCCTGTAGTCCAGCTACTCAGGAGGCTCAGGTGGAGGATTTGTTAGGCCAGA	38231		

Qy	1925	AC	TTTGGAGTCTGTGGCGGAGGATCACTTGAGCCCAAGGAGTTTGAGACTAGTCTGGGC	1984
Db	38057	GC	TTTGGAGGCCAAGTTGGGAGGATTCCTTGAGCCAGGAGTTCAAGACAGCCTGGGC	38116
Qy	1985	AACAGAGACA	-----TGTCTCTATAGTTGTGTTTGGTTTGTGTTTACCAAGTGTGGTG	2039
Db	38117	AACATAGCAAGACCCGTCTGTACAAAAAGTAAAAA	AAAAAAAAAAGTCAGCGCGTAG	38176
Qy	2040	TGTGCACCTGCAGTCCCGACTACTAGGAGGCGTGAAGTGGGAGGACTGCCCTGAGGCCAAG	2099	
Db	38177	TGTGCGCCTGTAGTCCAGCTACTCAGAGGCTGAGTGGGAGGATTCGTTTGAGCCCA	38236	
Qy	2100	AGGTCAGGCGTGAGTGAGCCATGATTTGTCACCTCGCACTCCAGCGCTGGGCAACACAGCA	2159	
Db	38237	AGGTCAAAGTGCAGTGAAGCAAGATTTGTGCCACTGTATCCAGCGCTGGGCAACAAAGCA	38296	
Qy	2160	AGACCTTGTCTCAAAAACAAACAAAGGCATCTCATAAAGTGCTCGGCTCCTATATGAT	2219	
Db	38297	AGACTGTGTCTCAAAAAAATAAGGGGGGGTTATTTCTGTGTTAATATATCTCTATATA	38356	
Qy	2220	TCAATATGT	2228	
Db	38357	TTAACATTT	38365	

```

RESULT 12
US-09-949-016-11909
; Sequence 11909, Application us/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11909
; LENGTH: 43255
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11909

```

Db	13162	GCACCTGCACCTCCAGCCTGGATGCAGAGCCAGACATATGTCTCAAAAAA	13222
Qy	1823	TTTTAAAAAAAT-AGGGGTACTAATATCTACCTTTAAAGGATGAGGGTTAAATTTAAGTAC	1881
Db	13222	AATTAATACACATGATGGAGCAAAACAGCTATGTAAACGCCTTTATTTTCAAAAGAAAA	13281
Qy	1882	ACACATAAGCCCTAGCGCAGTGGCTTATGCCCTGTAACTCAACACTTTGGGAGTCTGTGG	1941
Db	13282	ATACAGGTGGCCAGGCGAAGTTTCTCACGCCCTGTAAATGCCAGCACTTTTGGGAGGCTGAGG	13341
Qy	1942	CGGGAGGATCACTTGAGCCCGAGGAGTTTGAGACTAGTCTGGGCCAACAGAGACATGTCTCT	2001
Db	13342	CGGTGGATCGCTTAAGTTCAGGAGTTTGAGCAGCCAGGCTAACATGGTGAACCTCA	13401
Qy	2002	ATAGTTGTCTTTGGTTTTTTTTTACAGGTGTGGTGTGTGCACTGTGCAGTCCAGCTA	2061
Db	13402	TCCTACTAAAAATACAAAAATTAGCCGATGAGGTGGTGTGCTGTAGTCCAGCTA	13461
Qy	2062	CTAGGGAGGCTCAGGTGGGAGACCTGCTGAGCCAGGAGCTCGAGGCTCGAGTGAGCCA	2121
Db	13462	CTGGGAGGCTCAGGCGAGGAATCGCTTGAGCTTGGGAGGCGGAGGATGATGAGGCA	13521
Qy	2122	TGATTTGTCACCTGCACCTCCAGCCTGGGCAACACAGCAAGACCTTGTCTCAAAACAAAC	2181
Db	13522	AGATCAGCCATTGCACTCCAGCCTCGGTGCAGAGTGAGACTCGTCTTAAAAAACAA	13581
Qy	2182	AAAAA 2186	
Db	13582	ACAAA 13586	

RESULT 13

US-09-949-016-12453/c

; Sequence 1453, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12453

; LENGTH: 112219

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(112219)

; OTHER INFORMATION: n = A,T,C or G

; US-09-949-016-12453

[illegible]

RESULT 14

```

US-09-949-016-14324/c
; Sequence 14324, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14324
; LENGTH: 112222
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112222)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14324

```

## RESULT 15

```

RES011 13
US-09-949-016-17572/c
; Sequence 17572, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17572
; LENGTH: 113186
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17572

```

Query Match		13.0%	Score 322.2;	DB 4;	Length 113186;
Best Local Similarity		70.9%	Pred. No. 3.8e-58;		
Matches 487; Conservative		0;	Mismatches 188;	Indels 12;	Gaps 4;
Qy	1520	ACTGACGAGCGGGTGGCTCGGCGCTGTAATCCAGCACTTGGGAGGCCGAGGCAGCC	1579		
Db	65326	AATGGCGGGCACGGTGGCTCAGCCTGTAATCCAGCACTTGGGAGGCCGAGGCAGCC	65267		
Qy	1580	GGATCACTTGAGTCAAGATTTCAGACCGCCTGGTCAACATGGGTGAAACGCCGGCTC	1639		
Db	65266	GGATCAC--GAGGTCAGGAGATCGAGACCATCTCGCTAACAC--CAGTGAAACCCCGCTC	65210		
Qy	1640	TACCAAAATATAAAACTTAGCTGAGTGTGGTAGGCATGACTGTAATCCAGCAACTC	1699		
Db	65209	TACTAAAAACAAAAAAATAGCCGGCGTGGTGGGGCGCTGTAGTCCCGACTACTT	65150		
Qy	1700	AGGAGCTGAGGCA--GAGAAATCGCTTGAACCTGGGAGCGGAGGTTGCAGTGTGAGCTGAGA	1758		
Db	65149	GGGAGGCTGAGGCAGGAGAAATGCGGTGAACCCGGGAGGCAGAGCTTGCAGTGAGCCGAGA	65090		
Qy	1759	TCGTACACTGCACTCGAGCTTGGGGCAGAGCGGAGACTCTGCGCTTAAAAATAAATAAA	1818		
Db	65089	TCGCACCACTGCACTCGAGCTTGGGGCAGAGCGGAGACTCTCTCCAAAAAATAAAAAA	65030		
Qy	1819	TAATTTTAAAAAATAAGGGTACTAATAATCTACCTTAAAGGATGAGGTTAAATTAAG	1878		
Db	65029	AAAAAAGAGAAAAAATGAGAAACAGACTTCTTC-----TGGGATATGTTCAAAA	64978		
Qy	1879	TACACATAAGCCCTAGCGAGTGGCTTATGCTGTATCTCAACACTTTGGGAGTCTG	1938		
Db	64977	ATGTTATTCAGGCGCGGCACGGTGGCTCAGCGCTGTATCCCAATGCTTTGGAAGTCG	64918		
Qy	1939	TGGCGGAGGATCACTTGAGCCCGAGAGTTTGAGACTAGTCTGGGCAACAGAGACATGTC	1998		
Db	64917	CGGCAGGTGGATCACTTGAGGCCAGCTGTTCAAGAGCCAGCCTAGGCAACATGGTGAACC	64858		
Qy	1999	TCTATAGTTGTGTTTGGTTTGTGTTTACAGGTGGTGGTGTGCACCTGCAGTCCCAG	2058		
Db	64857	AGGTCTCTACTAAAAATACAAAAATAGCCAGGTGGTGGTGTACATGCCCTGTAAATCCCAG	64798		
Qy	2059	CTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCGAGGAGGTGAGGCTGCAGTGAG	2118		
Db	64797	CTACTTGGGAGGCTGAAATGGGAGGATCGCTTGAGCCCGAGGAGGCCAAGGTTGCAGTGAG	64738		
Qy	2119	CCATGATTGTGCCACTGCACTCCAGCTCGGCAACACAGCAAGACCTTGTCTCAAAAAACA	2178		
Db	64737	CCAAGATCCCACCACTGCACTCCAGCCTGAGTGACAGAGTAAGACCCCTGTCTTTAAAAAA	64678		
Qy	2179	AACAAAAAGCATACTCATAAAGTCTC	2205		
Db	64677	AAAAAAAAAAAAAAAAAAATATTTC	64651		

Search completed: August 12, 2005, 05:00:21  
Job time : 420 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 21:51:58 ; Search time 1296 Seconds  
(without alignments)  
11277.655 Million cell updates/sec

Title: US-10-829-118-2

Perfect score: 2469

Sequence: 1 ttgagctgtctggaggctc.....gaatacatggccctacagct 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2002bs:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2003ds:\*

12: Geneseqn2004as:\*

13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	6 AAD24181	Aad24181 Human sit
2	980	39.7	980	6 AAD24180	Aad24180 Human sit
3	333.8	13.5	4124	4 AAS01154	Aas01154 5'-untran
4	333.8	13.5	32145	4 AAK68575	Aak68575 Human imm
5	333.8	13.5	32145	4 AAK68491	Aak68491 Human imm
6	332	13.4	110000	13 ABD312780_2	Continuation (3 of
7	330.8	13.4	100445	13 ABD33179	Abd33179 Human can
8	327.4	13.3	147620	10 ADL113739	Adl113739 Osteoarth
9	327.4	13.3	147620	12 ADQ119948	Adq119948 Human sof
10	325.2	13.2	1307	8 ABZ69896	Abz69896 Human end
11	324.8	13.2	110000	12 ADQ97960_2	Continuation (3 of
12	324.2	13.1	110000	11 ACN44150_0	Acn44150 Human gen
13	322.4	13.0	105413	12 ADI36512	Adi36512 Human kin
14	321.6	13.0	91760	11 ACN44410	Acn44410 Human gen
15	321.4	13.0	47219	13 ADR16283	Adr16283 Human Cay
16	321.4	13.0	59884	13 ADR16284	Adr16284 Human Cay
17	319.2	12.9	94719	10 ADE95902	Ade95902 Human STA
18	319.2	12.9	94720	9 ADA02654	Ada02654 Human STA
19	319.2	12.9	94720	10 ADB72392	Adb72392 Human STA
20	318	12.9	24345	13 ADS36503	Ads36503 Human aut

C 21	318	12.9	40947	13	ADS36474	Ad36474 Human aut
C 22	317.8	12.9	192427	10	ADL13825	Adl13825 Osteoarth
C 23	317.2	12.8	270150	11	ADP65796	Adp65796 Human 16p
C 24	316.8	12.8	93390	10	ADD71350	Add71350 Glutamine
C 25	315.8	12.8	25236	4	AAK79879	Aak79879 Human imm
C 26	315.6	12.8	170834	10	AD62833	Ad62833 Human BAC
C 27	314.4	12.7	109906	6	ABK94411	Abk94411 DNA encod
C 28	314.4	12.7	109906	12	ADL08112	Adl08112 Human gen
C 29	314	12.7	5540	4	AAK79344	Aak79344 Human imm
C 30	313.8	12.7	133893	9	AAD54538	Aad54538 Human pho
C 31	313	12.7	304905	11	ADP75180	Adp75180 Human End
C 32	312	12.6	10770	4	AAK79181	Aak79181 Human imm
C 33	312	12.6	10770	4	AAS32561	Aas32561 Human gen
C 34	312	12.6	12613	4	AAK79178	Aak79178 Human imm
C 35	312	12.6	12613	4	AAS32560	Aas32560 Human gen
C 36	312	12.6	87687	11	ACN45166	Acn45166 Human gen
C 37	311.6	12.6	32249	4	AAL05336	Aal05336 Human rep
C 38	311.6	12.6	32249	4	ABL98205	AbL98205 Human tes
C 39	310.8	12.6	21721	2	AAK83427	Aak83427 Human 1p
C 40	310.8	12.6	39287	6	ABN80533	Abn80533 Human P45
C 41	310.4	12.6	2241	8	ABZ74035	Abz74035 Secrete
C 42	310.4	12.6	2241	8	ADA98642	Ada98642 Human sec
C 43	310.4	12.6	2241	10	ADC20765	Adc20765 Human sec
C 44	310.4	12.6	2241	10	ABZ67622	Abz67622 Human sec
C 45	310.2	12.6	22976	2	AAK83426	Aak83426 Genomic f

ALIGNMENTS

RESULT 1

AAD24181  
ID AAD24181 standard; DNA; 2469 BP.

XX AAD24181;

XX 07-MAY-2002 (first entry)

XX Human site-1 protease promoter region.

XX Site-1 protease; S1P; promoter; transcription; obesity; type II diabetes;  
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;  
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.

OS Homo sapiens.

XX WO200200873-A1.

XX 03-JAN-2002.

XX 19-JUN-2001; 2001WO-SE001386.

XX 27-JUN-2000; 2000SE-00002417.

XX (BIOV-) BIOVITRUM AB.

XX Abrahamsen L, Ekblom J, Forsgren M, Hoerling J, Johansson P;

XX WPI; 2002-139918/18.

XX Human site-1 protease promoter region for identifying agents capable of  
PT inhibiting the promoter activity useful in treating medical conditions  
PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia.

XX Claim 1; Page 15-16; 36pp; English.

XX The invention relates to human site-1 protease (S1P) promoter region. The  
CC promoter sequence is useful for identification of compounds that inhibit  
CC transcription of S1P, which in turn results in inhibition of sterol  
CC regulatory element-binding protein (SREBP) pathway. The compound  
CC identified is useful for the treatment of medical conditions related to  
CC obesity, type II diabetes, hypercholesterolemia, dyslipidaemia,  
CC atherosclerosis and other cardiovascular diseases. The present sequence



CC	is human SLP promoter region									
XX										
SQ	Sequence 2469 BP; 541 A; 704 C; 719 G; 505 T; 0 U; 0 Other;									
	Query Match 100.0%; Score 2469; DB 6; Length 2469;									
	Best Local Similarity 100.0%; Pred. No. 0;									
	Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	TTGAGTCTGTCTGGAGGCTCCGGCCAGAGCAGGGCGTATTGTTTCACTCGGTGAATGCT	60							
DB	1	TTGAGTCTGTCTGGAGGCTCCGGCCAGAGCAGGGCGTATTGTTTCACTCGGTGAATGCT	60							
QY	61	CATTTTCACTGTAAGAAAGAAACAGGCAACAGGAAACAAGCTGCGGAGCGCGAGACCCCGCA	120							
DB	61	CATTTTCACTGTAAGAAAGAAACAGGCAACAGGAAACAAGCTGCGGAGCGCGAGACCCCGCA	120							
QY	121	GGGCCCGGTACAGGCAACGCTGTGTCCAAACAAAGCCCGGAGCCCGCGCCACCTCC	180							
DB	121	GGGCCCGGTACAGGCAACGCTGTGTCCAAACAAAGCCCGGAGCCCGCGCCACCTCC	180							
QY	181	CCGACCCGGCCCGCCCGCCAGCCCTCGCTCGGGCCCTCGAGACGAACCCGGCACCT	240							
DB	181	CCGACCCGGCCCGCCCGCCAGCCCTCGCTCGGGCCCTCGAGACGAACCCGGCACCT	240							
QY	241	GAGCGAGCGGCGCCACCGCTAGCGGAGCGGGTTCGGGAGGCCCGCGCGCGGGCTG	300							
DB	241	GAGCGAGCGGCGCCACCGCTAGCGGAGCGGGTTCGGGAGGCCCGCGCGCGGGCTG	300							
QY	301	ACGTACTCTGCGCCCGGGAGCTCAGGGCGGGCCCGGGATGACGGCGCTCCGGCG	360							
DB	301	ACGTACTCTGCGCCCGGGAGCTCAGGGCGGGCCCGGGATGACGGCGCTCCGGCG	360							
QY	361	CGAACAGCGCTGGGCACTCAATTCGGGGCTGTTTACTCCCACTCTCGGAGACTGGGG	420							
DB	361	CGAACAGCGCTGGGCACTCAATTCGGGGCTGTTTACTCCCACTCTCGGAGACTGGGG	420							
QY	421	GCCGGCCAGCGAGGCCACAGCTGGAGGCTCAGCTCAGCTCCGCGACCCAGCTGCTGTC	480							
DB	421	GCCGGCCAGCGAGGCCACAGCTGGAGGCTCAGCTCAGCTCCGCGACCCAGCTGCTGTC	480							
QY	481	TGTCCTGCTCCCGGGCTTGGTGGCGCTCTGGAACGCTGGGAGCGGACCAACGC	540							
DB	481	TGTCCTGCTCCCGGGCTTGGTGGCGCTCTGGAACGCTGGGAGCGGACCAACGC	540							
QY	541	CGGGAGATGACGAAGAGTGTCTCGGACATTTTCGCGCGGGGGCCGGTGGCAGGGTGG	600							
DB	541	CGGGAGATGACGAAGAGTGTCTCGGACATTTTCGCGCGGGGGCCGGTGGCAGGGTGG	600							
QY	601	AAGCGAGGGCGTGGCCAGAGCTGCGAGGCGGAGAAACGCTGGGGAAACCTTTG	660							
DB	601	AAGCGAGGGCGTGGCCAGAGCTGCGAGGCGGAGAAACGCTGGGGAAACCTTTG	660							
QY	661	GTCCGCTCTGCGCGTCTAGGATCCCGAAAGAGACACGGGCGCGAAAGCGGCCAG	720							
DB	661	GTCCGCTCTGCGCGTCTAGGATCCCGAAAGAGACACGGGCGCGAAAGCGGCCAG	720							
QY	721	GCTGGGCGAGATCTAGAAAGACTGCTGGCGAGGCTCCCTGCCCGCGGGCTGCTG	780							
DB	721	GCTGGGCGAGATCTAGAAAGACTGCTGGCGAGGCTCCCTGCCCGCGGGCTGCTG	780							
QY	781	TCATGAGCTCGTAGAGGCTCGCTCCCGCGGACCTCTTCGCAAGGGTCCACGTCCA	840							
DB	781	TCATGAGCTCGTAGAGGCTCGCTCCCGCGGACCTCTTCGCAAGGGTCCACGTCCA	840							
QY	841	GGCACCGGGCTCGGACACCCACCCCGCGCGGACCTGCGCTGGGTGCCCTTTAAC	900							
DB	841	GGCACCGGGCTCGGACACCCACCCCGCGCGGACCTGCGCTGGGTGCCCTTTAAC	900							
QY	901	CCGGCGGTAGCTCGTTAAGATGGGAAAGTGTCCGGTCCGGAAACCGCGAAACCCCAAT	960							
DB	901	CCGGCGGTAGCTCGTTAAGATGGGAAAGTGTCCGGTCCGGAAACCGCGAAACCCCAAT	960							
QY	961	CCCGCTGCTCCGACCTCTGTACCCCGGCCCAACCGGACGACAGACTGGGCTCCCGAG	1020							
DB	961	CCCGCTGCTCCGACCTCTGTACCCCGGCCCAACCGGACGACAGACTGGGCTCCCGAG	1020							

Db 2041 GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGTGGAGGACTGCCTGAGCCAGGA 2100  
 QY 2101 GGTGAGGCTGCAGTGCAGCATGATTTGTCACCTGCACCTCCAGCCTGGGCAACACAGCAA 2160  
 Db 2101 GGTGAGGCTGCAGTGCAGCATGATTTGTCACCTGCACCTCCAGCCTGGGCAACACAGCAA 2160  
 QY 2161 GACCTTTGCTCAAAAAACAAACAAAAGCATACTCATAAAGTGTCTGGCTCCTATATGATT 2220  
 Db 2161 GACCTTTGCTCAAAAAACAAACAAAAGCATACTCATAAAGTGTCTGGCTCCTATATGATT 2220  
 QY 2221 CAATATGTTGGTGGTGAATCTTGAAATCTTCAATCTTCTGACTCAGATCTCATACGATTTTCTGA 2280  
 Db 2221 CAATATGTTGGTGGTGAATCTTGAAATCTTCTGACTCAGATCTCATACGATTTTCTGA 2280  
 QY 2281 ACTTTTGGAGATCTTGCCTCTCTGCAATTTGCAACCGTCAAGGACTCTCCCTCTGCC 2340  
 Db 2281 ACTTTTGGAGATCTTGCCTCTCTGCAATTTGCAACCGTCAAGGACTCTCCCTCTGCC 2340  
 QY 2341 ACCACACAAAGCATTTGATTTTAAACTTGAATATGTCTTCTTCTTCAACTTTAGGTAAA 2400  
 Db 2341 ACCACACAAAGCATTTGATTTTAAACTTGAATATGTCTTCTTCTTCAACTTTAGGTAAA 2400  
 QY 2401 TTAATCTTGGTCAAGGTTCTCTGAAACAGCCCTTTAGTCACTATGCAATGAATACATGGC 2460  
 Db 2401 TTAATCTTGGTCAAGGTTCTCTGAAACAGCCCTTTAGTCACTATGCAATGAATACATGGC 2460  
 QY 2461 CCTACAGCT 2469  
 Db 2461 CCTACAGCT 2469

## RESULT 2

AAD24180  
 ID AAD24180 standard; DNA; 980 BP.  
 XX AC  
 XX AAD24180;  
 DT 07-MAY-2002 (first entry)  
 DE Human site-1 protease promoter partial sequence.  
 XX  
 KW Site-1 protease; S1P; promoter; transcription; obesity; type II diabetes;  
 KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;  
 KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200200873-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 19-JUN-2001; 2001WO-SR001386.  
 XX  
 PR 27-JUN-2000; 2000SE-00002417.  
 XX  
 PA (BIOV-) BIOVITRUM AB.  
 XX  
 PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;  
 DR WPI; 2002-139918/18.  
 XX  
 PT Human site-1 protease promoter region for identifying agents capable of  
 PT inhibiting the promoter activity useful in treating medical conditions  
 PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia.  
 XX  
 PS Example 1; Page 15; 36pp; English.  
 XX  
 CC The invention relates to human site-1 protease (S1P) promoter region. The  
 CC promoter sequence is useful for identification of compounds that inhibit  
 CC transcription of S1P, which in turn results in inhibition of sterol  
 CC regulatory element-binding protein (SREBP) pathway. The compound  
 CC identified is useful for the treatment of medical conditions related to  
 CC obesity, type II diabetes, hypercholesterolaemia, dyslipidaemia,

CC atherosclerosis and other cardiovascular diseases. The present sequence  
 CC is a fragment of human S1P promoter  
 XX  
 SQ Sequence 980 BP; 175 A; 313 C; 326 G; 166 T; 0 U; 0 Other;  
 Query Match 39.7%; Score 980; DB 6; Length 980;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-207;  
 Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 353 CTCGCGGCGAACAACGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCACTCTCCGAG 412  
 Db *Seq 1* 1 CTCGCGGCGAACAACGCTGGGCACTCCATTTCGGGGCTGTTTACTCCCACTCTCCGAG 60  
 QY 413 ACTGGGCGCGCGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 472  
 Db 61 ACTGGGCGCGCGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
 QY 473 GGCCTGTGTCTCCCGGCTTCCCGGGCTTTCGCTGGCGCTCTGGACGCGTGGGCGAGCGG 532  
 Db 121 GGCCTGTGTCTCCCGGCTTCCCGGGCTTTCGCTGGCGCTCTGGACGCGTGGGCGAGCGG 180  
 QY 533 GACCACGCGGAGGATGAGCAAGTGTCTCGGACATTTGCGGCGGCGGCGGCGGCGGCGG 592  
 Db 181 GACCACGCGGAGGATGAGCAAGTGTCTCGGACATTTGCGGCGGCGGCGGCGGCGGCGG 240  
 QY 593 CAGGGTGGAAAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 652  
 Db 241 CAGGGTGGAAAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 QY 653 AACCTTGTCTCCGCTCTGCGGCTCTAGGATCCCGAAAGAGAGACGCGGCGGCGGAA 712  
 Db 301 AACCTTGTCTCCGCTCTGCGGCTCTAGGATCCCGAAAGAGAGACGCGGCGGCGGAA 360  
 QY 713 GCGGCGAGGCTGGGCGGAGGATCTAGAAAGACTGCTTCCCGCGGCGGCGGCGGCGGCGG 772  
 Db 361 GCGGCGAGGCTGGGCGGAGGATCTAGAAAGACTGCTTCCCGCGGCGGCGGCGGCGGCGG 420  
 QY 773 GCTGTGTCTGAGACTCGTGGAGAGCTCGCTTCCCGCGGCGGCGGCGGCGGCGGCGGTC 832  
 Db 421 GCTGTGTCTGAGACTCGTGGAGAGCTCGCTTCCCGCGGCGGCGGCGGCGGCGGTC 480  
 QY 833 CAGCTCCAGGCAACGCGGCTCGGACACCCACCCCGGCGGCGGCGGCGGCGGCGGCGGTC 892  
 Db 481 CAGCTCCAGGCAACGCGGCTCGGACACCCACCCCGGCGGCGGCGGCGGCGGCGGCGGTC 540  
 QY 893 CCCTTAACCCGGGCGGTAGCTGTTAAGATGGGAAGTGTCCGCTCCGGAACACGCGAA 952  
 Db 541 CCCTTAACCCGGGCGGTAGCTGTTAAGATGGGAAGTGTCCGCTCCGGAACACGCGAA 600  
 QY 953 CCCCAATCCCGCTCGCCGACCTCTGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGTC 1012  
 Db 601 CCCCAATCCCGCTCGCCGACCTCTGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGTC 660  
 QY 1013 TCCGACGCGGCGGCGGCTGCGGGGACACCGGTGCTGCGGAAACGAGGAGCTTTGTAAC 1072  
 Db 661 TCCGACGCGGCGGCGGCTGCGGGGACACCGGTGCTGCGGAAACGAGGAGCTTTGTAAC 720  
 QY 1073 GCCACGTGTTGCTCTTTTGAAGAAACAGATTAATGTTTAAACTGTCGAAAAGCT 1132  
 Db 721 GCCACGTGTTGCTCTTTTGAAGAAACAGATTAATGTTTAAACTGTCGAAAAGCT 780  
 QY 1133 TGCCGCTTAAAGATGCTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAA 1192  
 Db 781 TGCCGCTTAAAGATGCTCTGGGTGACTTAGATGCTAGGATCAGTTTGTCTTCAATGTAAA 840  
 QY 1193 TGGACCAAGCCCGGACTCCGTAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1252  
 Db 841 TGGACCAAGCCCGGACTCCGTAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 QY 1253 TGGTGGGCGGTGATGCGCTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1312  
 Db 901 TGGTGGGCGGTGATGCGCTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960

QY 1313 CTGCACCGGTAATGCTGA 1332  
 |||||  
 Db 961 CTGCACCGGTAATGCTGA 980

RESULT 3  
 AA01154/c  
 ID AA01154 standard; DNA; 4124 BP.  
 XX AA01154;  
 XX  
 DT 12-SBP-2001 (first entry)  
 XX  
 DE 5'-untranslated region interferon induced nucleic acid, IFN3.  
 XX Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;  
 KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;  
 KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;  
 KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;  
 KW haematologic disease; chronic neutropenia; myocardial infarction;  
 KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;  
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN3; ds.  
 XX Homo sapiens.  
 XX  
 PH Location/Qualifiers  
 FT Promoter  
 FT 404...411  
 FT /tag= a  
 FT /note= "Gamma-interferon activation site"  
 FT 499...506  
 FT /tag= b  
 FT /note= "Gamma-interferon activation site"  
 FT 891...898  
 FT /tag= c  
 FT /note= "Gamma-interferon activation site"  
 FT 1519...1526  
 FT /tag= d  
 FT /note= "Gamma-interferon activation site"  
 FT 1653...1660  
 FT /tag= e  
 FT /note= "Gamma-interferon activation site"  
 FT 1687...1694  
 FT /tag= f  
 FT /note= "Gamma-interferon activation site"  
 FT 2918...2930  
 FT /tag= g  
 FT /note= "Interferon-stimulated response element"  
 FT 3708...3715  
 FT /tag= h  
 FT /note= "Gamma-interferon activation site"  
 XX WO200118208-A2.  
 XX  
 PD 15-MAR-2001.  
 XX  
 XX 08-SEP-2000; 2000WO-US024704.  
 XX  
 PR 08-SEP-1999; 99US-0152921P.  
 PR 20-OCT-1999; 99US-0160575P.  
 PR 20-JAN-2000; 2000US-0177104P.  
 PR 07-SEP-2000; 2000US-00656633.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 PA (BIOJ) BIOGEN INC.  
 XX  
 XX Payman JA, Da Silva A, Hochman P, Hau A;  
 XX WPI; 2001-235201/24.  
 XX  
 XX New interferon induced polypeptides and polynucleotides, useful for the  
 PT diagnosis, prevention and treatment of immunological, cell proliferative  
 PT disorders, such as lupus erythematosus, cancer, stroke and Alzheimer's  
 PT disease.

XX Disclosure; Page 27-29; 134pp; English.  
 PS  
 XX The sequence represents the 5'-untranslated region of interferon induced  
 CC nucleic acid, IFN3. IFN nucleic acids and polypeptides are useful for  
 CC treating or preventing a pathology associated with IFN polypeptide in a  
 CC human. They are useful for determining the presence of or predisposition  
 CC to a disease associated with altered levels of IFN polypeptide or  
 CC polynucleotide. IFN nucleic acids, polypeptides and antibodies are useful  
 CC for diagnosis, prevention or treatment of variety of immunological and  
 CC cell proliferative disorders, such as autoimmune diseases e.g. lupus  
 CC erythematosus, immunodeficiency diseases such as acquired  
 CC immunodeficiency syndrome (AIDS), graft rejection, viral infections  
 CC including hepatitis and human immunodeficiency virus (HIV), immune-  
 CC mediated glomerulonephritis, haematologic diseases such as aplastic  
 CC anaemia and chronic neutropenia and cancer. In addition they are also  
 CC useful for treating or diagnosing various disorders associated with cell  
 CC death, including myocardial infarction, stroke, neurological diseases  
 CC including Alzheimer's and Parkinson's diseases, amyotrophic lateral  
 CC sclerosis and spinal muscular atrophy. IFN nucleic acids and polypeptides  
 CC are also useful for identifying interferon-like proteins and interferon  
 CC agonists, for screening drugs and compounds which inhibit or enhance IFN  
 CC activity or function and as targets for the identification of small  
 CC molecules that are immunostimulatory, immunosuppressive, or stimulate or  
 CC suppress normal cell or tumour cell growth in mammals, including humans  
 XX  
 XX Sequence 4124 BP; 1167 A; 873 C; 878 G; 1206 T; 0 U; 0 Other;  
 SQ

Query Match 13.5%; Score 333.8; DB 4; Length 4124;  
 Best Local Similarity 66.4%; Pred. No. 6.3e-64;  
 Matches 525; Conservative 0; Mismatches 262; Indels 4; Gaps 3;

QY 1503 ATTACAAAATAGGGGTACTAGCCAGGGGGGTGCTCGCGCTCTAATCCAGCACTTG 1562  
 |||||  
 Db 1264 ATTGAAAAAATAAATTTCAAGCCAGGCACATGGCTCAGCGCTGTAATGCTAGCATT 1205  
 |||||  
 QY 1563 GGGAGCCGAGGAGCGCGGATCATTGAGGTGAGAAGTTTCAGACAGCGCTGGTCAACAT 1622  
 |||||  
 Db 1204 GGGAGCCAGGTGGTGGATCCTGAGTTAGAGTTTGAGACATCTGGTCCACA 1145  
 |||||  
 QY 1623 GGGTGAACCGCGGCTTACCAAAAATATAAAAACTTAGCTGAGTGTGTAGCGCATGAC 1682  
 |||||  
 Db 1144 TAGTGAACCCCTGTCTCTACTATAAACACGAAAA-TTAGCTGGCATGGTGGCAGCACC 1086  
 |||||  
 QY 1683 TGTATCCAGCAACTCAGGAGCTGAGGCA-GAGAACTGCTTGAACCTGGAGGGGAG 1741  
 |||||  
 Db 1085 TGTATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATGCTTGAACCTGGAGGTGGAG 1026  
 |||||  
 QY 1742 GTTGCACTGAGCTGAGATCGTACCCTGCACTCCAGCTTGGCGGACAGAGCGAGACTCTG 1801  
 |||||  
 Db 1025 GTTGCACTGAGACCGAGATCGCACCTGCACTCCAGCTTGGCGGACAGAGCGAGACTCGG 966  
 |||||  
 QY 1802 CCTTAAAAATAAATAAATTTTAAAAAATAAGGGGTACTAATAATCTACTTAAAGG 1861  
 |||||  
 Db 965 TCTCAAAATAAATAAATAAATAAATAAATAAATTTTCATCTGTTCTCTTAGCTTTAAA 906  
 |||||  
 QY 1862 ATGAGGTTAATTAAGTACACACATAAGCCCTAGCGGAGTGGCTTATGCCCTGTAATCTC 1921  
 |||||  
 Db 905 ATGATGTTTCTAGAAAAATAAATAAAC--TAGGCGGGCGGTGGCTCAGCGCTGTAATCCC 848  
 |||||  
 QY 1922 AACACTTTGGGAGTCTGTGGCGGAGGATCACTTCAGCCCGAGAGTTTGAGACTAGTCTG 1981  
 |||||  
 Db 847 AGCACTTTGGGAGGCGGAGGAGGCGGATCAGAGGTGAGGAGATCAGACCATCTTGGC 788  
 |||||  
 QY 1982 GGCAACAGAGACATGTCTCTATAGTTGTGTTTGGTTTTTTTACAGGTGTGGTGGTG 2041  
 |||||  
 Db 787 TAACACGGTGAACCCCGTCTCTACTGAAAAATACAAAAAATTAGCCGGCGGTGGTGGTG 728  
 |||||  
 QY 2042 TGCACCTGAGTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTCCCTGAGCCAGGAG 2101  
 |||||  
 Db 727 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATGGCGTGAACCCGGGAG 668  
 |||||  
 QY 2102 GTCGAGGCTGCAGTGCAGCCATGATTGTGCCACTGCACTCCAGCTGGGCAACACAGCAAG 2161  
 |||||

Db 667 GCGGAGCTTCAGTGCAGCAGATCGCGCACTGCACCTCCAGCGCTCGGCGGAAAGAGCGAG 608  
Qy 2162 ACCTCTGCTCAAAAAACAACAAAAAGCATACTCATAAAGTGTCTGGCTCCTATATGATTC 2221  
Db 607 ACTCTGCTCAAAAAACAACAAAAAATAAATAAAAAAATAAAACTTATGCTCTTGGCTTACAT 548  
Qy 2222 AATATGTGGTGGTGATCTTTGAATCCTTTCCTGACTCAGATCTCATACGATTTTCTGAA 2281  
Db 547 TGTTCTTATTGGATGCTGCTCTACACCAAGCGTCATTTCCTTTCTGAAACTCTATCACT 488  
Qy 2282 CTTTGGAGAA 2292  
Db 487 GTTTTGTAA 477

RESULT 4  
AAK68575/c  
ID AAK68575 standard; DNA; 32145 BP.  
XX  
XX AAK68575;  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23387.  
DE  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226581P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.



PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-02114886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 26-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246521P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246603P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-483426/52.  
 Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 useful for preventing, diagnosing and/or treating cancers and metastasis.  
 Disclosure; SEQ ID NO 23303; 3071pp + Sequence Listing; English.  
 AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 activity, and can be used in gene therapy and vaccine production. (I)



CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 U; 0 Other;

		Query Match	13.5%;	Score 333.8;	DB 4;	Length 32145;		
		Best Local Similarity	66.4%;	Pred. No. 1.1e-63;				
		Matches	525;	Conservative	0;	Mismatches 262;	Indels	4; Gaps 3;
QY	1503	ATTACAAAATAGGGTACTAGCAGCGGGGGTGGCTCGCGCTGTAAATCCCGACACTTG	1562					
DB	3586	ATTGAAAAATAAATTTCAAGGCAGGCACAGTGGCTCAGCGCTGTAAATGCTAGCACTTT	3527					
QY	1563	GGGAGCGGAGCGGCGGATCACTTGAGTCAAGATTTTCAGACGAGCTGGTCAACAT	1622					
DB	3526	GGGAGCGGAGCGGCTGATCACTTGAGTCAAGATTTTCAGACGAGCTGGTCAACAT	3467					
QY	1623	GGGTGAAACCGCGCTCTACCAAAAATATAAAAATTTAGCTGAGTGGTAGCGCATGAC	1682					
DB	3466	TAGTGAACCCCTGCTCTACTAAAAACACGAAAA--TTAGCTGGGCATGGTGGCAGGCACC	3408					
QY	1683	TGTAAATCCCGAACAACTCAGAGGCTGAGGCA--GAGAATCGTTTGAACCTGGGAGCGGAG	1741					
DB	3407	TGTAAATCCCGAGTACTTGGAGGCTGAGGCGAGGAATCGTTGAACCTGGGAGGTGGAG	3348					
QY	1742	GTTCAGTGGAGTGTGATCGTACCTGCACTCCAGCTTGGGCGACAGAGCGAGACTCTG	1801					
DB	3347	GTTCAGTGGAGTGTGATCGTACCTGCACTCCAGCTTGGGCGACAGAGCGAGACTCTG	3288					
QY	1802	CCTTAAAAATAAATAAATTTTTTAAAAAAAATAGGGTACTTAATCTACCTTAAAGG	1861					
DB	3287	TCTCAAAATAAATAAATAAATAAATAAATAAATTTATCTGTTCTTCTTACGTTTTAA	3228					
QY	1862	ATGAGGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTATCTC	1921					
DB	3227	ATGATGTTTCTAGAAAAATAAAAC--TAGGCGGGGGTGGTGGCTCAGCGCTGTAATCCC	3170					
QY	1922	AACACTTTGGAGTCTGTGGCGGAGGATCACTTGAGCCACAGAGTTTGAGACTAGTCTG	1981					
DB	3169	AGCACTTTGGAGGCGGAGGACGCGGATCACAGGTCAGGAGTCAAGACCATCTCTGGC	3110					
QY	1982	GGCAACAGACATGCTCTATAGTGTGTGTTGTTGTTTGTGTTTACAGGTGGTGGTG	2041					
DB	3109	TAACACGGTGAACCCCGTCTCTACTGAAATAACAAAAATTTAGCGGGCGTGGTGGTG	3050					
QY	2042	TGCACCTGCACTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCCTGAGCCGAGGAG	2101					
DB	3049	GGCACCTGTAATCCAGCTACTTTGGAGGCTGAGGCGAGGAGATGGCGTGAACCCGGAG	2990					
QY	2102	GTGAGGCTGAGTGGAGGATGATTTGGCACTGCACTCCAGCTGGGCGAACACAGCAAG	2161					
DB	2989	GCGGAGCTTGCAGTGAAGAGATTCGCGCACCTGCACCTCCAGCTGGGCGAAAGAGCGAG	2930					
QY	2162	ACCTGTCTCAAAAACAAACAAAAGCATACTATAAGTGTCTCGGCTCTATATGATTC	2221					
DB	2929	ACTCTGTCTCAAAAACAAACAAAATATAAATAAATAAATAAATAAATAAATAAATAAATA	2870					
QY	2222	AATATGTGGTGGTGTCTTGAATCTTTCTGACTCAGATCTCATACGATTTCTGAA	2281					
DB	2869	TGTTTCTATTGGATGCTGCTCTACACCAACCGTCAATTTCCCTTTCTGAAACTCTACACT	2810					

QY	2282	CTTTTGGAGAA	2292					
DB	2809	GTTTTGTATA	2799					
RESULT 6								
ABD32780_2/c								
Continuation (3 of 5) of ABD32780 from base 200001 (Human cancer-associated genomic DNA)								
WP Sequence split into 5 fragments LOCUS ABD32780 Accession ABD32780								
WP	Fragment Name	Begin	End					
WP	ABD32780_0	1	110000					
WP	ABD32780_1	100001	210000					
WP	ABD32780_2	200001	310000					
WP	ABD32780_3	300001	410000					
WP	ABD32780_4	400001	430442					

		Query Match	13.4%;	Score 332;	DB 13;	Length 110000;		
		Best Local Similarity	70.8%;	Pred. No. 4e-63;				
		Matches	487;	Conservative	0;	Mismatches 190;	Indels	11; Gaps 3;
QY	1508	AAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAAATCCCGACACTTGGGGAG	1567					
DB	59424	AAATATTCAGAAAGAGGCCAGCGCGGTGTTTCATGCTGTAAACCCCGACACTTTGGGAG	59365					
QY	1568	GCGAGGCGAGCGGATCACTTCAGGTCAAGTTCAGACACAGCCTGGTCAACATGGGTG	1627					
DB	59364	GCTGAGTGGGCGGATGACCTAAGGTGAGCATTTGAGACGAGCTGGCCACAT--GGTG	59306					
QY	1628	AAACCGCGCTCTACCAAAAATATAAAATTTAGTGTAGTGGTGGTACGATGATGATAA	1687					
DB	59305	AAACCGCTGTCTACTTAAAAATACAAAAAGTTAGCTGAGCATGGTGGCTCAGATCTGATA	59246					
QY	1688	TCCGAGCACTCAGGAGGCTGAGGCA--GAGATCGTTCAACCTGGGAGCGGAGGTGC	1746					
DB	59245	TCCGAGTACTCGGAGGGTGGAGGAGGAATCGCTTGAACCCAGAGCAGAGGTTC	59186					
QY	1747	AGTGAGTCAAGATCGTACCACTGCACCTCCAGCTTGGGCGACAGAGCAGACTCTGCCTTA	1806					
DB	59185	AGTGAGCAGATCGACCACTTGCATTCAGCTTGGGCGACAGAGCAGACTCCGTTCTCA	59126					
QY	1807	AAAAATAAATAAATTTTTTAAAAAAAATAGGGGTACTTAATATCTTAAAGGATGAG	1866					
DB	59125	AAAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	59066					
QY	1867	GGTTAAATTA-----AGTACACACATAGCCCTAGCGAGTGGCTTATGCTGTAA	1917					
DB	59065	GCAAGCATTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA	59006					
QY	1918	TCTCAACACTTTGGGAGTCTGTGGGCGGAGGATCACTTGAGCCGAGGAGTTTGAGACTAG	1977					
DB	59005	TCCGAGCTTCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	58946					
QY	1978	TCTGGGCAACAGAGACATGCTCTATAGTGTGTGTTGTTGTTTGTGTTTACAGGTGTGT	2037					
DB	58945	CCTGGCCAACTGGTGAACCTGTCTACAAAAATAAATAAATAAATAAATAAATAAATAAATA	58886					
QY	2038	GGTGTGCACTGAGTCCGAGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCA	2097					
DB	58885	GGTGTGCTGCTATAATATCCAGCTACTTTGGGAGACTGAGGAGGAGGAGGAGGAGGAGG	58826					
QY	2098	GGAGGTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2157					
DB	58825	AGGATGAGAGGTTGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG	58766					
QY	2158	CAAGACCTTGTCTCAAAAACAAACAAAA	2185					
DB	58765	TGAGACCTTGTCAAAAACAAACAAAAA	58738					

RESULT 7  
 ABD33179/c  
 ID ABD33179 standard; DNA; 100445 BP.  
 XX





CC associated with a disease, preferably osteoarthritis. The cell line and  
 CC the non-human animal are useful for screening for an agent for diagnosing  
 CC an individual having susceptibility to joint space narrowing and/or  
 CC osteophyte development and/or joint pain. This sequence corresponds to  
 CC the polynucleotide encoding a protein listed in the specification. (Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences).

XX Sequence 147620 BP; 39001 A; 34978 C; 35560 G; 38081 T; 0 U; 0 Other;

Query Match 13.3%; Score 327.4; DB 10; Length 147620;  
 Best Local Similarity 71.0%; Pred. No. 4.6e-62;  
 Matches 489; Conservative 0; Mismatches 196; Indels 4; Gaps 4;

QY 1504 TTACAAAATAAGGGGTACTAGCAGCGGGGTGGCTGCGCGCTGTAAATCCAGCACTTGG 1563  
 Db 40080 TCAAAAATAATAAATAAGCGCGGGTGTGGTGGCTCAGCGCTGTAAATCTCAGCACTTGG 40139  
 QY 1564 GGAGCGGAGCGCGGATCACTTGGCTGAGGCTTTCAGACCGCTGGTCAACATG 1623  
 Db 40140 GGAGCGGAGCGCGGATCACTTGGCTGAGGCTTTCAGACCGCTGGTCAACAT- 40198  
 QY 1624 GGTGAAACGCCGCTCTACCAAAAATATAAAAACTTAGCTGAGTGTGGTAGCGCATGACT 1683  
 Db 40199 GGTGAAACGCCGCTCTACTAAAAATACAAAAA-TTAGCTGGCGGTGGTGGTGGCT 40257  
 QY 1684 GTAAATCCAGCAACTCAGGAGGCTGAGGCA-GAGAAATCGCTTGAACCTGGAGCGGAGG 1742  
 Db 40258 CTAAATCCAGCACTTGGGAGGCTGAGGCAAGAAATCACTTGAATCCAGAGCGGAGG 40317  
 QY 1743 TTGCACTGAGTCACTGCTGACCACTGCACTCCAGCTTGGGACACAGCGGAGACTCTGC 1802  
 Db 40318 TTGCACTGAGCTAGATCATGCCACTGCTATCCAGCTGGGTGATAGAGCGAGACTCAGT 40377  
 QY 1803 CTTAAAAATAAATAAATAAATTTTAAAAAATAAGGGGTACTTAATCTACCTTAAAGGA 1862  
 Db 40378 CGAAGAAG 40437  
 QY 1863 TGAGGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCA 1922  
 Db 40438 AGAAGAAAAG 40496  
 QY 1923 ACATTTGGGAGTCTGTGGCGGAGGATCACTTGGCCAGGAGTGTGAGACTGTCTGG 1982  
 Db 40497 ACATTTGGGAGGCTGAGATGGGTGGATCACTTGGGTCAAGAGTTCGAGACCGAGCTGG 40556  
 QY 1983 GCAACAGAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTGTGTTTGTGTTTGTGTT 2042  
 Db 40557 CGAATGTTGAAACCCGCTCTACTAAAAATACCAAAATCAGCCAGCATGTGTAGCAT 40616  
 QY 2043 GCATCTGAGTCCAGCTACTAGGAGGCTGAGGTGGAGGAGTGTGCTGAGCCCGAGAGG 2102  
 Db 40617 GCATCTGTAATCCAGCTACTTGGAGGCTGATGTCAGAGGAATCACTTGAATCTGGAGG 40676  
 QY 2103 TCGAGCTGAGTGGAGGATGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2162  
 Db 40677 CGAGGTTTGAGTGGAGGAGATCAAGCCATTTGATCTCCAGCTGGGCAACCGAGGAGGA 40736  
 QY 2163 CTTTGTCTCAAAAACAAAACAAAAGGATA 2191  
 Db 40737 CTTTAAACAAAAG 40765

RESULT 9  
 ADQ19948/c  
 ID ADQ19948 standard; DNA; 147620 BP.

XX AC ADQ19948;

XX XX 26-AUG-2004 (first entry)

XX DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 2768.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
 KW da.  
 KW Homo sapiens.  
 OS WO2004048938-A2.  
 PN 10-JUN-2004.  
 PD 26-NOV-2003; 2003WO-US038193.  
 PF 26-NOV-2002; 2002US-0429739P.  
 PR (PROT-) PROTEIN DESIGN LABS INC.  
 PA Aziz N, Ginsburg WM, Zlotnik A;  
 PI WPI; 2004-441208/41.  
 DR Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 XX Example 2; SEQ ID NO 2768; 210pp; English.  
 PS The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples, where a higher level  
 CC of expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC DNA of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 XX Sequence 147620 BP; 38081 A; 35560 C; 34978 G; 39001 T; 0 U; 0 Other;

Query Match 13.3%; Score 327.4; DB 12; Length 147620;  
 Best Local Similarity 71.0%; Pred. No. 4.6e-62;  
 Matches 489; Conservative 0; Mismatches 196; Indels 4; Gaps 4;

QY 1504 TTACAAAATAAGGGGTACTAGCAGCGGGGTGGCTGCGCGCTGTAAATCCAGCACTTGG 1563  
 Db 107541 TCAAAAAATAAATAAATAAGCGCGGGTGTGGTGGCTCAGCGCTGTAAATCTCAGCACTTGG 107482  
 QY 1564 GGAGCGGAGGCGCGGATCACTTGGAGTCAAGAGTTTTCAGACCGCTGGTCAACATG 1623  
 Db 107481 GGAGCGGAGGCGCGGATCACTTGGAGTCAAGAGTTTTCAGACCGCTGGTCAACAT- 107423  
 QY 1624 GGTGAAACGCCGCTCTACCAAAAATATAAAAACTTAGCTGAGTGTGGTAGCGCATGACT 1683  
 Db 107422 GGTGAAACGCCGCTCTACTAAAAATACAAAAA-TTAGCTGGCGGTGGTGGTGGCT 107364  
 QY 1684 GTAATCCAGCAACTCAGGAGGCTGAGGCA-GAGAAATCCCTTTGAACTCCGGAGCGGAGG 1742  
 Db 107363 CTAATCCAGCTACTTGGAGGCTGAGGCAAGAAATCACTTGAATCCAGAGGCGGAGG 107304  
 QY 1743 TTGCACTGAGTCACTGAGTCACTGCACTCCAGCTTGGCGACAGAGCGAGACTCTGC 1802  
 Db 107303 TTGCACTGAGCTAGTCACTGCACTTCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 107244  
 QY 1803 CTTAAAAATAAATAAATAAATTTTAAAAAATAAGGGGTACTTAATCTACCTTAAAGGA 1862  
 Db 107243 CGAAGAAG 107184  
 QY 1863 TGAGGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCA 1922  
 Db 107184 TGAGGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGCTGTAAATCTCA 1922

Db 107183 AGAAGAAAAAGAAAGAGA-AGAAACAGCCAGGTGTCAGTGGCTCATGCTGTAAATCCA 107125  
QY 1923 ACACCTTTGGAGTCTGTGCGGAGGATCACTTGGCCAGGAGTTTGAGACTAGTCTGG 1982  
Db 107124 ACACCTTTGGAGCTGAGATGGTGTGATCCTTGGTTCAGAGTTCAGACAGCCTGG 107065  
QY 1983 GCAACAGAGACATGCTCTATAGTTGTGTTGGTTTGTGTTTACAGGTGGTGGTGT 2042  
Db 107064 CCAACATGTTGAACCCCGTCTCTACTAAAAATACCAAAATCAGCCAGACATGTTAGCAT 107005  
QY 2043 GCACCTGACGTCCTCAGCTACTAGGAGGCTGAGGTGGAGGAGTCTGCTGAGCCAGGAGG 2102  
Db 107004 GCATCTGTAATCCAGCTACTTTGGGAGGCTGATGCAGGAGAAATCACTTGAACCTTGGGAGG 106945  
QY 2103 TCGAGGCTCAGTGAGCCATGTTGTCACCTGCACCTCCAGCTGGGCAACACAGCAAGA 2162  
Db 106944 CGGAGTTTGAGTGAGCCGAGATCAGCCCATTTGATCTCCAGCTGGGCAACGAGGAGGA 106885  
QY 2163 CTTTGTCTCAAAAAACAAACAAAAAGCATA 2191  
Db 106884 CTCATAAACAAAAAGGAAAAAGAAA 106856

RESULT 10  
ABZ69896/c  
ID ABZ69896 standard; cDNA; 1307 BP.  
XX AC ABZ69896;  
XX DT 23-APR-2003 (first entry)  
XX DE Human endothelial cell differentiation factor protein 9.35 cDNA.  
XX KW Human; endothelial cell differentiation factor protein 9.35; EDP-1;  
XX KW cancer; HIV; gene; ss.  
XX OS Homo sapiens.  
XX FT Key Location/Qualifiers  
XX CDS 351..608  
XX FT /\*tag= a  
XX FT /product= "EDF-1 9.35"

XX CN1363571-A.  
XX PD 14-AUG-2002.  
XX 05-JAN-2001; 2001CN-00105043.  
XX 05-JAN-2001; 2001CN-00105043.  
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX MAO Y, XIE Y;  
XX WPI; 2003-000312/01.  
XX P-PSDB; ABP1850.  
XX Polypeptide-human endothelial cell differentiation factor EDF-1 protein  
XX 9.35 and polynucleotide for coding it.  
XX Claim 6; Page 25-26 (Disclosure); 32pp; Chinese.  
XX The invention relates to the novel human endothelial cell differentiation  
XX factor (EDF-1) protein 9.35, and the polynucleotide encoding it. The  
XX polypeptide is useful in treating diseases such as cancer and HIV  
XX infection. The antagonist of the polypeptide and its medical function,  
XX and the application of the polynucleotide are also disclosed. The present  
XX sequence encodes the human EDF-1 protein 9.35 of the invention  
XX  
XX Sequence 1307 BP; 285 A; 327 C; 323 G; 372 T; 0 U; 0 Other;  
XX Query Match 13.2%; Score 325.2; DB 8; Length 1307;

Best Local Similarity 71.0%; Pred. No. 3.7e-62;  
Matches 487; Conservative 0; Mismatches 193; Indels 6; Gaps 4;  
QY 1506 ACAAAATAGGGTACTAGCCAGCGGGTGGCTCGCCCTGTATCCAGCAGCTTGGG 1565  
Db 1070 AAATATTTATAAACACAGCCAGCGGTGGTTCAGCCCTGTATTCAGCAGCTTGGG 1011  
QY 1566 AGCCGAGCGCAGCCGATCACTTGAAGTTCAGAGTTTCAGACCCGCTTCAACATGGG 1625  
Db 1010 AGCCCAAGCGCGGAGATCACTGAGGTTCAGTAGTTCGAGACCCCTAGCTAAACAT-GG 952  
QY 1626 TGAACCGCGGCTCTACCAAAATATAAAATTAAGTTCAGTGTGTCGAGCAGCTGCT 1685  
Db 951 TGAACCCCGCTCTCACTTAAATAACAAAAATTCATCGGGTGTGGCGCGCGCTGT 892  
QY 1686 AATCCAGCACTCAGGAGGCTGAGGCA-GAATAATCGCTTGAACCTGGGAGCGGAGGTT 1744  
Db 891 AATCCAGCTACACGGAGCTGAGCAGGAGATCGCTTAAGCCCGGAGAGCGAGGTT 832  
QY 1745 GCAGTGAGTGAATCGTACCACTGCACTCCAGCTTGGGCGACAGAGGAGACTCTGCT 1804  
Db 831 GCAGTGAGTGAATCGCGCACTGCACTCCAGCTGGGCGACAAAGCGAGACTCTGTCT 772  
QY 1805 TAAAAATTAATAATTTTAAAAAATAGGGTACTATATATCTTACCTTAAAGGATG 1864  
Db 771 TAAAAAATAATATATATATATATATATATATATATATCTCTACAAAAATTCCTAGATCTC 712  
QY 1865 AGGGTTAAATTAAGTACACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTATCTCAAC 1924  
Db 711 TGACCACAAAAATCCGCCCCCTAGCTGGGCGGGTGGCTCAGCCCTGTATCCAGC 652  
QY 1925 ACTTTGGAGTCTGTGGGAGGATCACTTGAAGCCAGGAGTTTGAAGACTAGTCTGGGC 1984  
Db 651 ACTTTGGAGCGGAGTGGTGGATCACTCAGGTTCAGAGTTCAGACAGCAGCTTGGCC 592  
QY 1985 AACAGAGATCTCTATAGTTGTGTTGTTTGTGTTTGTGTTTGA-CCAGGTGTGGTGTG 2043  
Db 591 AACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATCAGCCGGGTGTGGTGGCGG 532  
QY 2044 CACCTGCACTCCAGCTACTAGGAGGCTGAGTGGGAGGCTGCTGAGCCGCTTGAAGCC 2100  
Db 531 CGCTGTATATCCAGCTACACGGAGGCTGAGCAGGAGATCGCTTGAACCCGGAGGA 472  
QY 2101 GGTGAGGCTGCACTGAGCCATGATGTGCTCCACTGCACTCCAGCTTGGGCAACACAGCA 2160  
Db 471 GACAGAGTTGCACTGAGCCAAAGATCGCCCATTCGACTCCAGCTTGGGCGACAGAGCGA 412  
QY 2161 GACCTTGTCTCAAAACAAACAAAAA 2186  
Db 411 AACACCGTCTTAAAAAATAAAAAA 386

RESULT 11  
ADQ97960.2/c  
Continuation (3 of 4) of ADQ97960 from base 200001 (Human cancer associated sequence HDL1  
WP Sequence split into 4 fragments LOCUS ADQ97960 Accession Adq97960  
WP Fragment Name Begin End  
WP ADQ97960\_0 110000  
WP ADQ97960\_1 100001 210000  
WP ADQ97960\_2 200001 310000  
WP ADQ97960\_3 300001 390183  
Query Match 13.2%; Score 324.8; DB 12; Length 110000;  
Best Local Similarity 68.6%; Pred. No. 1.6e-61;  
Matches 528; Conservative 0; Mismatches 222; Indels 20; Gaps 5;  
QY 1524 GCAGCGCGGGTGGCTCGCCCTGTATCCAGCAGTTCGGGAGGCGGAGGAGCCGGAT 1583  
Db 15258 GCCAGCGCGGTGGCTCAGCCCTGTATCCAGCAGCTTTGGGAGGCTGAGGCGGGGGAT 15199  
QY 1584 CACTTGAAGTCAAGATTTTCAGACCCAGCTTGTCAACATGGGTGAACCGCGCTTACC 1643  
Db 15198 CA--TAAGGTTCAGGAGATCAAGACCGCTGGCTAAACA-CGATGAACCTCCTTCTACT 15142

QY 1644 AAAAAATATAAATCTAGCTGAGTGTGTAGCGCATGATGTAAATCCCAAGCACTCAGGA 1703  
 Db 15141 AACAAATACAAAAAATAGCTGGCGGTGGTAGCACAATGCTGTAGTCTCAGCTACTCAGGA 15082  
 QY 1704 GGCTGAGGCA-GAGAAATCGCTTGAACCTGGAGCGGAGGTGACGTGAGCTGAGATCGT 1762  
 Db 15081 GGCTGAGGCAAGAGATCACTTGAACCTGAGAGCGGAGCTTGCAGTGAGCAAGATCG 15022  
 QY 1763 ACACATGCACTCCAGCTTGGCGGACAGAGCGAGACTCTGCTTTAAAAATAAATAAT 1822  
 Db 15021 GCCACTGCACTCCAGCTGGCGGACAGGCAAGACTCCATTTAAAAAATAAATAATCTT 14962  
 QY 1823 TTTTAAAAAATAGGGTACTAATAATCTACCTTTAAAGGATGAGGTTAAATTAAGTACA 1882  
 Db 14961 ACCTCTTCAATTAATAATCTTAACAGCAGTAATAAAGATGAGG----- 14917  
 QY 1883 CACATAAGCCCTAGCGCAGTGGCTTATGCTGTAAATCTCAACACTTTGGAGTCTGTGGC 1942  
 Db 14916 CAGATCTGGCGGCGCAGTGGCTCATGCTGTAAATCCCACTTTGGAGGCTGAGGT 14857  
 QY 1943 GGAGATGACCTTGAGCCAGGAGTTGAGACTAGTCTGGGCAACAGAGACATGTCTCTA 2002  
 Db 14856 GGCGGATGACCTGAGCTCAGGAGTTGAGACCAGCTGGCCAACTGTGTGAAACCTGT 14797  
 QY 2003 TAGTTGTGTTGGTTTTTTTACAGGTTGGTGTGTCAGCTGCACTCCAGCTAC 2062  
 Db 14796 CTCTACTAAAAATACAAAAATAGCTGGGTGGCGGCGCTGTAAATCCCACTAC 14737  
 QY 2063 TAGGAGGCTGAGTGGGAGGACTGCTGAGCCAGGAGTCCAGGCTGAGTGGCCAT 2122  
 Db 14736 TTGGGAGATTGAGCGAGGAGATCGCTTGAACCCAGGAGCGAGGTTGCACTGAGCTGA 14677  
 QY 2123 GATTGTGCCACTGCACTCCAGCTT-GGGCAACACAGCAAGACCTTGTCTCAAAAAACAAC 2181  
 Db 14676 GATCAGGCCACTGCACTCCAGCTGGGGGAAAGAGCGAGCTTTGTCTCAAAAAACAAC 14617  
 QY 2182 AAAAAATCTACTAATAGTCTGGCTCTCTATGATGATCAATGATGATGATGATGATGAT 2241  
 Db 14616 AATGAGGCGAGATCTACATGAACTGACACAGAAATAATCAATTAAGATGTGCTGGTAGTGA 14557  
 QY 2242 TGAATCTTTCTCTGACTCAGATCTCATACGATTTTCTGAACTTTTGGAGA 2291  
 Db 14556 ABAATGCAATAGCACAAAGTGTGTATGCTTATTTTAAATTTTTTAAA 14507

RESULT 12  
 ACN44150\_0  
 WP Sequence split into 5 fragments LOCUS ACN44150 Accession Acn44150  
 WP Fragment Name Begin End  
 WP ACN44150\_0 1 110000  
 WP ACN44150\_1 100001 210000  
 WP ACN44150\_2 200001 310000  
 WP ACN44150\_3 300001 410000  
 WP ACN44150\_4 400001 439892  
 ID ACN44150 standard; DNA; 439892 BP.  
 XX ACN44150;  
 AC ACN44150;  
 XX 18-NOV-2004 (first entry)  
 DT Human genomic sequence hCG27278.  
 DE  
 XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003073826-A2.  
 XX  
 PD 12-SEP-2003.  
 XX  
 PF 28-FEB-2003; 2003WO-US006235.  
 XX

PR 01-MAR-2002; 2002US-00087192.  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX Morris DW;  
 PI  
 XX WPI; 2003-328604/31.  
 DR  
 XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 PT comprises a nucleotide sequence.  
 PT  
 XX Claim 1; SEQ ID NO 454; Opp; English.  
 PS  
 XX The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent to basic patent  
 CC US2002182586A1, for which no sequence data was published  
 CC  
 XX Sequence 439892 BP; 100791A; 10435C; 106090G; 116168T; 0U; 124080Other;  
 SQ  
 Query Match 13.1%; Score 324.4; DB 11; Length 110000;  
 Best Local Similarity 71.2%; Pred. No. 1.9e-61;  
 Matches 500; Conservative 0; Mismatches 191; Indels 11; Gaps 5;  
 QY 1497 TTCTTCATTACAAAATAGGGGTACTAGCCAGCGGGGTGGCTCGCGCTGTAATCCCAAG 1556  
 Db 22994 TTTCTAAGTCTAAATATATGGAATAATGGCCAGCACAGTGGCTCAGCGCTGTAATCCCAAG 23053  
 QY 1557 GACTTGGGAGCCGAGCAGCGGATCACTTGAAGTCAAGAGTTTCAGACAGCCTGGT 1616  
 Db 23054 CACTTGGAGGCTCAGGAGGCGAGATCACTTGAAGTCAAGAGTTTGAACAGCCTGGC 23113  
 QY 1617 CAACATGGGTGAAACGCCGCTCTACCAAAAAATATAAAAACTTAGCTAGTGTGGTAGCG 1676  
 Db 23114 CACATGTGTAAACCTGTTTCTACTAAAAATCCAAAA-TCAGCCGGCATGGTGGCA 23172  
 QY 1677 CATGACTGTAATCCCAAGCAACTCAGAGGCTCAGGCA-GAGAAATCGTTGAACCTGGAG 1735  
 Db 23173 GATGCTGTAAATCCCAAGCTACTCGGAGGCTCAGGCGAGGAGAAATCACTTGAACCCAGGAG 23232  
 QY 1736 GCGGAGTTGCACTGAGCTGAGATCGTACCAGTCCAGCTCCAGCTTGGGCGAGCAGAGCGAG 1795  
 Db 23233 GCAGAGTTTGCAGTGAAGCGGATCGTCCACTGCACTCCAGCTCCAGCTGGGCGAGCAGAGCGAG 23292  
 QY 1796 ACTCTGCCTTAAAAATAAATAAATAATTTTAA-AAAAAATAGGGGTACTAATAATCTACC 1854  
 Db 23293 ACTCGCTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 23352  
 QY 1855 TTAAGGATGAGGTTAAATTAAGTACACACA-----TAAGCCCTAGCGAGTGGCTT 1907  
 Db 23353 TAAATATTGGCATTTGCAAAATCAATGCTTTAACTGTGTTCTAGGCTGGGCGACAGTGGCTG 23412  
 QY 1908 ATGCTGTAAATCTCAACACTTTGGGAGTCTGTGGGAGGAGTCACTTTCAGAGCCAGGAGT 1967  
 Db 23413 ATGCTGTCAATCCAGCACTTTTGGGAGGCTTGGGAGGCTTGGGAGGATCATTTGAGCTCAGGAT 23472  
 QY 1968 TTGAGACTAGTCTGGGCAA-CAGAGACATGTCTCTATAGTGTGTTGTTGTTTGT 2026  
 Db 23473 TTGAGACCAAGTGGCAATGTGGTGAACCTCTGTCTTCAAAAAATAAATAAATAAATAAATAA 23532  
 QY 2027 CCAGGTGTGGTGTGCACTCCAGTCCAGTCTAGGAGGCTGAGGTTGGAGGAGT 2086  
 Db 23533 CCGGGCATGGTGGCATGTGCTGTGCTGCCAGCTACTCGGATGGCTGAGTCAAGGGGATG 23592



```

FT      /*tag= ap
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(52308,C)
FT      /tag= aq
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(52333,T)
FT      /tag= ar
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(53079,C)
FT      /tag= as
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      exon      59182..59445
FT      /tag= at
FT      intron      59446..88092
FT      /tag= au
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(62141,T)
FT      /tag= av
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(70338..70340,CT)
FT      /tag= aw
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(70379,A)
FT      /tag= ax
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(76074,T)
FT      /tag= ay
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(77336,C)
FT      /tag= az
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(77582,C)
FT      /tag= ba
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(78010,C, T)
FT      /tag= bb
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(78011,G, C)
FT      /tag= bc
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(79557,G)
FT      /tag= bd
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(79617,C)
FT      /tag= be
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(84900,G)
FT      /tag= bf
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      exon      88093..88188
FT      /tag= bg
FT      intron      88189..93291
FT      /tag= bh
FT      exon      93292..93349
FT      /tag= bi
FT      intron      93350..98719
FT      /tag= bj
FT      exon      98720..98797
FT      /tag= bk
FT      intron      98798..102286
FT      /tag= bl
FT      variation  replace(100529,G)
FT      /tag= bm
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation  replace(100996..100998,CC)
FT      /tag= bn
FT      /standard_name= "Single nucleotide polymorphism (SNP)"

```

```

Query Match      13.0%; Score 322.2; DB 12; Length 105413;
Best Local Similarity 70.9%; Pred. No. 5.9e-61;
Matches 487; Conservative 0; Mismatches 188; Indels 12; Gaps 4;

```

```

Db      64357 AATGGCGGGCACGGTGGCTCACGGCTGTAATCCACGACACTCTGGAGAGCCGAGCGGGC 64298
Qy      1580 GGATCACTTGAGGTGAGAGTTTCAGACGAGCTGGTCAACATGGGTGAAACGCCGGCTC 1639
Db      64297 GGATCAC--GAGGTGAGGAGATCGAGACCATCTGGCTAACA-CAGTGAACCCCGCTC 64241
Qy      1640 TACCAAAAATAAAAACTTAGCTGAGTGGCATGACTGTAAATCCCAAGCAACTC 1699
Db      64240 TACTAAAAACAAAAAATTAGCCGGCGTGGTGGGGCTGTAGTCCAGCTACTT 64181
Qy      1700 AGGAGCTGAGCA-GAGAAATCGTTGAACCTGGGAGCGGAGGTTCAGTGGAGTGAGA 1758
Db      64180 GGGAGGCTGAGGCAGAGAAATGGCGTGAACCCGGGAGGAGAGCTTGCAGTGAGCGAGA 64121
Qy      1759 TCGTACCCTGCACTCCAGCTTGGGGCAGAGCGAGACTCTGCCTTAAAAATATAAA 1818
Db      64120 TCGCACCACTGCACTCCAGCTTGGGGCAGAGCGAGACTCTCTCCAAAAAATAAA 64061
Qy      1819 TAATTTTAAAAAATAAGGGTACTTAATATCTACCTTAAAGGATGAGGGTTAAATTAAG 1878
Db      64060 AAAAAAGAGAAAAATGAGAAAAGAGACTTTCTTC-----TGGGATATGTTCAAA 64009
Qy      1879 TACACACAATAGCCCTAGCGCAGTGGCTTATGCCCTGTAACTCAACAATTTGGAGTCTG 1938
Db      64008 ATGTTATTCAAGGCGGGCAGCGTGGCTCAGCCTGTAATCCCAATGCTTTGGAAGTGC 63949
Qy      1939 TGGCGGAGGATCACTTGGAGCCAGAGTTTGAGACTAGTCTGGGCAACAGAGACATGTC 1998
Db      63948 CGGCAGGTGGATCACTTGGAGCCAGCTGTTCAAGACCAGCCTAGGCAACATGGTGAACC 63889
Qy      1999 TCTATAGTTGTGTGTTGTTTGTACAGGTGGTGGTGTGCACCTGCAGTCCCAAG 2058
Db      63888 AGTCTCTACTAAAAATACAAAAATTAGCCAGGTGGTGGTACATGCTGTAAATCCAG 63829
Qy      2059 CTACTAGGAGGCTGAGGTGGAGGACTGCCCTGAGCCAGGAGGTGAGGCTGCACTGAG 2118
Db      63828 CTACTTGGAGGCTGAAATGGGAGGATCGCTTGAGCCAGGAGGCCAAGGTTGCACTGAG 63769
Qy      2119 CCATGATTTGTCACCTGCACTCGAGCTGGGCAACACAGCAAGACCTTGTCTCAAAAA 2178
Db      63768 CCAAGATCCCACTGCACTCCAGCTGAGTGACAGAGTAAGACCCCTGTCTTTAAAAA 63709
Qy      2179 AACAAAAAGCATACTCATAAAGTGCTC 2205
Db      63708 AAAAAAATAAAAAAATAAATTAATTC 63682

RESULT 14
ID ACN44410/c
XX ACN44410 standard; DNA; 91760 BP.
AC ACN44410;
XX ACN44410;
DT 18-NOV-2004 (first entry)
XX 18-NOV-2004 (first entry)
DE Human genomic sequence hCG41365.
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX Homo sapiens.
XX WO2003073826-A2.
XX 12-SEP-2003.
PD 28-FEB-2003; 2003WO-US006235.
XX 28-FEB-2003; 2003WO-US006235.
PR 01-MAR-2002; 2002US-00087192.
XX 01-MAR-2002; 2002US-00087192.
PA (SAGR-) SAGRES DISCOVERY.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX Morris DW;

```



DR WPI; 2003-328604/31.  
 XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 PT comprises a nucleotide sequence.  
 PS Claim 1; SEQ ID NO 844; Opp; English.  
 XX The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent to basic patent  
 CC US2002182586A1, for which no sequence data was published  
 XX SQ Sequence 91760 BP; 21177 A; 23589 C; 24129 G; 22845 T; 0 U; 20 Other;  
 Query Match 13.0%; Score 321.6; DB 11; Length 91760;  
 Best Local Similarity 72.1%; Pred. No. 7.7e-61;  
 Matches 490; Conservative 0; Mismatches 179; Indels 11; Gaps 5;  
 QY 1523 AGCCGCGGGTGGCTGCGCCCTGTAATCCAGCACTTGGGAGCGCCGAGCGCGGA 1582  
 Db 68876 AGCCGCGCGAGTGGCTGCGCCCTGTAATCCAGCACTTGGGAGCGCCGAGCGAGTGA 68817  
 QY 1583 TCACCTGAGTCAAGATTTTCAGACCCAGCTGTCTACATGGGTGAACCGCGCTCTAC 1642  
 Db 68816 TCAC--GAGGTCAAGCATTTTTCAGACCCAGCTGTCTACAT--GGTGAACCGCGCTCTAC 68760  
 QY 1643 CAAAAATATAAACTTAGCTGAGTGTGGTGGCGCATGCTGTAATCCAGCACTCAGG 1702  
 Db 68759 TAAATATACAAATATTAGCCAG---GCACGGTGGCAGCGCTTAACCCCACTACTCCGG 68703  
 QY 1703 AGGCTGAGGC-AGAGAAATCGCTTGAACCTGGGAGCGGAGGTTGCGATGAGCTGAGATCG 1761  
 Db 68702 AGGCTGATGCGGAGAACTACTTTGAACCGAGGAGGTGAAGGTTGCGATGAGCAAGACCG 68643  
 QY 1762 TACCACTGCACTCCAGCTTGGGCGACAGAGCGAGCTCTCCCTTAAATAAATAATAA 1821  
 Db 68642 TGCCATTGCACTCCAGCTTGGGCGACAGAGCGAGCTCTCTCTCAAAAAAATAAAAAA 68583  
 QY 1822 TTTTAAAAAATAAGGGGTACTAATATCTACCTTAAAGGATGAGGTTAAATTAAGTAC 1881  
 Db 68582 AAAAAAAGGAGAGGCTTGGACAGTGCACATAGTATGCCAGTTTGTGTTTAAAGA 68523  
 QY 1882 ACACATAAGCCCTAGCGAGTGGCTTATGCTGCTGTAATCTCAACACTTTGGAGTCTGTGG 1941  
 Db 68522 AAAGTGTAGGCTGGATGCAATGGCTCGCACCCTACAGTCCAGCACTTTGGGAGCGGAGG 68463  
 QY 1942 CGGAGGATCACTTGACCCAGGAGTTTGAGACTAGTCTGGGCAACA----GAGACATGT 1997  
 Db 68462 CAGAGGATTTGCTTTAGGCGCAGGAGTTCAAGACAGCGCTGGGCAACATAGCGAGAACCTG 68403  
 QY 1998 CTCTATATGTTGTTTGGTTTGTGTTTATACAGGTGGTGTGTGCACTGCGAGTCCCA 2057  
 Db 68402 TTCTATGAAGATTAATTAATAATTAAGCAGCGCTGGTGTGTGCACTGTAGTCTTCCCA 68343  
 QY 2058 GCTACTAGGAGGCTGAGGTGGGAGGATGCTGCTGAGCCCGAGGAGGTGAGGCTGCACTGA 2117  
 Db 68342 GCTACTTTGGGAGGCTGTGTGTAGGAGGATGGCTTCTAGTCTAGGAGTTTGGGCTGCACTGA 68283  
 QY 2118 GCGATGATTTGCCACTGCACTCCAGCTCGGCGCAACACAGCAGACACTTGTCTCAAAAAC 2177  
 Db 68282 GCTATGATGGCGCGCTGCACTCCAGCTTGGGCAACACAGTGAACCGCTGTCTCTAAAT 68223  
 QY 2178 AAACAAAAGCATACTCATATA 2197

Db 68222 AAAATAAAGGTTTTTTAA 68203  
 RESULT 15  
 ADRI6283/c  
 ID ADRI6283 standard; DNA; 47219 BP.  
 XX AC ADRI6283;  
 XX DT 21-OCT-2004 (first entry)  
 XX DE Human Cayman ataxia splice site mutant DNA.  
 XX KW Human; jittery; de; Cayman ataxia; ATCAV; KIAA1872; chromosome 19p13.3;  
 KW ataxia; myoclonus; dystonia; epilepsy; nyctagmus; splice site mutant;  
 KW SNP; single nucleotide polymorphism.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT misc\_feature 33181..33190  
 FT /tag= a  
 FT /note= "Exon 9-intron 9 boundary"  
 FT variation replace(33187,G)  
 FT /tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "Causes splice site mutation"  
 XX US2004146900-A1.  
 PN 29-JUL-2004.  
 XX PD  
 XX 03-NOV-2003; 2003US-00699941.  
 XX PF  
 XX 01-NOV-2002; 2002US-0422971P.  
 XX PR 08-NOV-2002; 2002US-0424973P.  
 XX PA (UNMI ) UNIV MICHIGAN.  
 XX PI Burmeister M;  
 XX WPI; 2004-552665/53.  
 XX DR  
 XX PT Detecting variant Cayman ataxia polypeptide or nucleic acid sequence in  
 PT subject, by detecting presence or absence of variant Cayman ataxia  
 PT polypeptide or nucleic acid in biological sample.  
 XX PS Claim 5; SEQ ID NO 10; 98pp; English.  
 XX CC The invention relates to detecting variant Cayman ataxia polypeptide or  
 CC nucleic acid sequence (ATCAV) in a subject, involving providing a  
 CC biological sample from a subject, where the biological sample comprises a  
 CC Cayman ataxia polypeptide or nucleic acid, and detecting the presence or  
 CC absence of a variant Cayman ataxia polypeptide or nucleic acid in the  
 CC biological sample. Also included is a kit comprising a reagent for  
 CC detecting the presence or absence of a variant Cayman ataxia nucleic acid  
 CC or polypeptide in a biological sample. The human ATCAV gene (previously  
 CC isolated as cDNA KIAA1872, located on chromosome 19p13.3) was identified  
 CC and mapped using the fact that it is the homologue of the mouse jittery  
 CC gene, located on mouse chromosome 10. The method is useful for detecting  
 CC variant Cayman ataxia polypeptide or nucleic acid sequence in a subject.  
 CC The presence of the variant Cayman ataxia polypeptide or nucleic acid is  
 CC indicative of a disorder chosen from ataxia, myoclonus, dystonia,  
 CC epilepsy, and nyctagmus in the subject. The biological sample is chosen  
 CC from blood sample, a tissue sample, urine sample, saliva sample, and an  
 CC amniotic fluid sample. The subject is chosen from embryo, fetus, newborn  
 CC animal, young animal, and an adult animal. The animal is a human. The  
 CC human is an adult female of child-bearing age. The present sequence is  
 CC the DNA representing a Human ATCAV splice site mutant (disease causing)  
 CC where a single nucleotide polymorphism in the splice site of exon 9-  
 CC intron 9 causes a read through mutant protein to be produced.  
 XX CC



8Q	Sequence	47219 BP; 11562 A; 12203 C; 12045 G; 11209 T; 0 U; 200 Other;
	Query Match	13.0%; Score 321.4; DB 13; Length 47219;
	Best Local Similarity	68.9%; Pred. No. 7.1e-61;
	Matches	513; Conservative 0; Mismatches 226; Indels 6; Gaps 5;
Qy	1461	TTTGTGACGAGTTGTTTAACTCTTTGTACCTCTTTGTACCTCAGCTCTTTCATTACAAAAATAGGGTA 1520
Db	20591	TCCTGTGCAGAGAATTGGNAATAGTTTATGTACAGTTCTGCCCTCAAGNAGATGAAGGA 20532
Qy	1521	CTAGCCAGCGGGTGGCTCGCCCTGTATATCCAGACCTTGGGAGGCGGAGCGAGCCG 1580
Db	20531	TAGCCCGGGTGCCTGCTCAGCCTGTATCCAGACCTTTGGAGGCGGAGCGTGGC 20472
Qy	1581	GATCCTTTGAGGTCAGAGTTTCAGACCAGCCTGGTCACATGGGTGMAACCGCGCTCT 1640
Db	20471	GATCAC--GAGGTCAGAGATCGAGACCACCTCTGGCTAACAT--GGTGAACCCCGTCTCT 20415
Qy	1641	ACCAAAAAATA-TAAAACTTAGCTGAGTGTGGTAGCGCATGACTGTAAATCCAGCAACTC 1699
Db	20414	ACTAAAAGTACAAAAAATAAGCGGGCGTGGTGCGAGTGCCTGTAGTCCAGCTACTC 20355
Qy	1700	AGGAGGCTGAGGCA-GAGATCGCTTGAACCTGGGAGGCGAGGTTGCAGTGAGCTGAGA 1758
Db	20354	AGGAGGCTGAGGCGAGGAATGCGATGAACCCAGGAGGCGAGCTTGCAGTGAGCGAGA 20295
Qy	1759	TGTTACCACTGCACCTCAGCTTGGGCGACAGAGCGGAGACTCTGCCTTAAAAATAATAA 1818
Db	20294	TCGTGCCACTGCACCTCAGCCTGGGCGCACAGACAGACTCTGTCTCAAAAAAATAAAA 20235
Qy	1819	TAATTTTAAAAAATAAGGGTACTAAATATCTACCTTAAAGGATGAGGGTTAAATTAAG 1878
Db	20234	AAAAAAAAGATGAGGATAACTCTGTGTAGGCTGTGTATGTATGTATGTATTCANAA 20175
Qy	1879	TACACACATAAGCCCTAGCGAGTGGCTTATGCTGTATCTCAACACTTTGGGAGTCTG 1938
Db	20174	AATACAGGATGGCCAGGCGCAGAGCTCAGGCTGTATATCCAGCACTTTGGGAGGCTG 20115
Qy	1939	TGGGGGAGGATCACTTGAGCCGAGAGTTTGAGACTAGTCTGGGCAACAGACATGTC 1998
Db	20114	AGGCAGGAGGATCACTTGAGCTCAGGAATTCGAGACCGCCTGGCAACATGTTGAAACG 20055
Qy	1999	TCTATAGTTGTGTTTGTGTTTGTGTTTACCAGGTGGTGGTGTGCACTGCACTCCAG 2058
Db	20054	CCATCTCTATAAATAACAAAAATTAGGCGAGCATGGTGGAGGACACCTGTATATCCAG 19995
Qy	2059	CTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCGAGGAGGTGAGGCTCAGTGAG 2118
Db	19994	CTA-TCAGGAGGCTGAGGCGAGGAGATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAG 19936
Qy	2119	CCATGATTGTGCCACTGCACCTCAGCCTGGGCAACACAGCAAGACCTTGTCTCAAAAACA 2178
Db	19935	CCGAGATCGTGCCACTGCTCTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAACA 19876
Qy	2179	AACAAAAAGCATACTCATAAAGTGC 2203
Db	19875	ACAAACAAACAAACTATCAAGGC 19851

Search completed: August 11, 2005, 23:45:38  
Job time : 2326 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 11, 2005, 22:35:08 ; Search time 10557 Seconds  
(without alignments)  
11332.379 Million cell updates/sec

Title: US-10-829-118-2  
Perfect score: 2469  
Sequence: 1 ttgagtcgtctcggaggtcc.....gaatacatggccctacagct 2469

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	6	AR527528 Sequence
2	2447.4	99.1	212596	9	AC040169 Homo sapi
3	2447.4	99.1	216522	2	AC139664 Homo sapi
C	980	39.7	980	6	AR527527 Sequence
4	521.4	21.1	532	9	BC052334 Homo sapi
C	347	14.1	148864	9	Z77249 Human DNA s
5	346	14.0	46741	9	AC010519 Homo sapi
C	344	13.9	194241	2	AC093324 Homo sapi
6	343.4	13.9	167587	9	AC108000 Homo sapi
C	339.4	13.7	52479	2	AC074125 Homo sapi
7	339.4	13.7	86684	9	AC010605 Homo sapi
C	339.4	13.7	169583	9	AC080005 Homo sapi
8	339	13.7	197992	2	AC020715 Homo sapi
C	338.4	13.7	173252	9	AC036176 Homo sapi
9	338.4	13.7	188923	9	AC009267 Homo sapi
C	338.2	13.7	231160	2	AC135717 Homo sapi
10	337	13.6	185624	9	AC103923 Homo sapi
C	336.4	13.6	43457	9	AL499628 Human DNA
11	336.4	13.6	81579	9	AE006640 Homo sapi

20	336.4	13.6	110000	2	BX324168_4	Continuation (5 of
21	335.8	13.6	109488	9	HS298J15	AL031774 Human DNA
22	335	13.6	158800	2	AC027090	AC027090 Homo sapi
23	335	13.6	192096	2	AL590133	AL590133 Human DNA
24	334.8	13.6	221357	2	AP004246	AP004246 Homo sapi
25	334.4	13.5	201659	2	AC146419	AC146419 Pan trogl
26	334.2	13.5	184393	9	AC018665	AC018665 Homo sapi
C	333.8	13.5	110293	9	HS298J15	AX098245 Sequence
27	333.8	13.5	110293	9	HS298J15	AL049766 Human DNA
28	333.8	13.5	110293	9	HS298J15	AL109965 Human DNA
C	332.8	13.5	160979	2	AC055879	AC055879 Homo sapi
29	332.8	13.5	211063	2	AC069254	AC069254 Homo sapi
C	332.2	13.4	97120	9	AP004147	AP004147 Homo sapi
30	332	13.4	349980	6	CQ869981	CQ869981 Sequence
C	331.4	13.4	50417	9	AY497474	AY497474 Homo sapi
31	331.4	13.4	101624	9	AC104116	AC104116 Homo sapi
C	331.4	13.4	160570	2	AC092317	AC092317 Homo sapi
32	330.8	13.4	157871	9	AC104365	AC104365 Homo sapi
C	330.6	13.4	123501	9	AL390241	AL390241 Human DNA
33	330.6	13.4	145166	2	AC053473	AC053473 Homo sapi
C	330.6	13.4	164810	2	AC023411	AC023411 Homo sapi
34	330.6	13.4	266079	2	AL391060	AL391060 Homo sapi
C	330.2	13.4	163338	9	AL162426	AL162426 Human DNA
35	330.2	13.4	172588	2	AC007430	AC007430 Homo sapi
C	330	13.4	33026	9	AF156673	AF156673 Homo sapi
36	330	13.4	177893	9	AC005089	AC005089 Homo sapi

ALIGNMENTS

RESULT 1  
AR527528  
LOCUS AR527528 2469 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 2 from patent US 6723553.  
ACCESSION AR527528  
VERSION AR527528.1 GI:53914633  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2469)  
AUTHORS Abrahamsen,L., Ekblom,J., Forsgren,M., Horling,J. and Johansson,P.  
TITLE Promoter sequences  
JOURNAL Patent: US 6723553-A 2 20-APR-2004;  
FEATURES  
source  
1..2469  
/organism="unknown"  
/mol\_type="genomic DNA"

Query Match	100.0%	Score 2469;	DB 6;	Length 2469;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Indels 0;
Matches 2469;	Conservative 0;	Gaps 0;		
Qy	1	TTGAGTCTCTCTGGAGGCTCCGGCCAGACGAGCGGTATTGTTTTCACTCGGTGAATGCT	60	
Db	1	TTGAGTCTCTCTGGAGGCTCCGGCCAGACGAGCGGTATTGTTTTCACTCGGTGAATGCT	60	
Qy	61	CATTTTCACTAAAGAAAAACAGGCAACGAAAGCTGCGGAGCGGAGACCCCGCA	120	
Db	61	CATTTTCACTAAAGAAAAACAGGCAACGAAAGCTGCGGAGCGGAGACCCCGCA	120	
Qy	121	GGGCGCGGTACAGGACCGCTGTGTCCAAACAGCGCCGGAGCCCGCCACCTCCC	180	
Db	121	GGGCGCGGTACAGGACCGCTGTGTCCAAACAGCGCCGGAGCCCGCCACCTCCC	180	
Qy	181	CGGACCGCGCCCGCCCGCCCGCTCGCTCGGGGCTCGGACGCAACCGCACACCT	240	
Db	181	CGGACCGCGCCCGCCCGCCCGCTCGCTCGGGGCTCGGACGCAACCGCACACCT	240	
Qy	241	GAGCGAGCGGCGCCACCGCTAGGCGGAGCGCGGTGCGGAGCGCGCGCGGCTG	300	

Db 241 GAGGAGCGGCGCCACCGCTAGGCGAGCGGGTCGGGGAGGCGCGCGCGCGGGCTG 300  
Qy 301 AGGTACTGCGCGCGCGGGAGCTCAGGGCCGGCGGGCCGGGATGACGGGGCCCTCCGGG 360  
Db 301 ACGTACTGCGCGCGCGGGAGCTCAGGGCCGGCGGGCCGGGATGACGGGGCCCTCCGGG 360  
Qy 361 CGAACACGCTGGGCACTCCATTCGGGGCTGTGTTACTCCCAACTCTCGCGAGACTGGGG 420  
Db 361 CGAACACGCTGGGCACTCCATTCGGGGCTGTGTTACTCCCAACTCTCGCGAGACTGGGG 420  
Qy 421 GCGGGCCAGGAGGCCACAGCTGGGAGCTCAGCTCGGCCACCCAGCGTGGCCCTGTG 480  
Db 421 GCGGGCCAGGAGGCCACAGCTGGGAGCTCAGCTCGGCCACCCAGCGTGGCCCTGTG 480  
Qy 481 TGTCCCGCTCCCGGGCTTGGTGGGGCTCTGGAACGCGGTGGGAGCGGGACCAACGC 540  
Db 481 TGTCCCGCTCCCGGGCTTGGTGGGGCTCTGGAACGCGGTGGGAGCGGGACCAACGC 540  
Qy 541 CGGGAGGATGGACGAAGGTGCTCGCACATTTGCGGCGGGGGGGCCGGTGGCAGGGTG 600  
Db 541 CGGGAGGATGGACGAAGGTGCTCGCACATTTGCGGCGGGGGGGCCGGTGGCAGGGTG 600  
Qy 601 AAGCGAGGGCGTGGCCAGGAGCTGCGAGCGGGAGAACGCGCTGGGGGAACCCCTTG 660  
Db 601 AAGCGAGGGCGTGGCCAGGAGCTGCGAGCGGGAGAACGCGCTGGGGGAACCCCTTG 660  
Qy 661 GTCGGCTGTCGGCTCGCTTAGGATCCCGGAAGAGGACCGGGCGCGAAGCGGCCAG 720  
Db 661 GTCGGCTGTCGGCTCGCTTAGGATCCCGGAAGAGGACCGGGCGCGAAGCGGCCAG 720  
Qy 721 GCTGGGCGAGGATAGAAAGCTGCTGCGCAGGCTCCCTGCCCGGGCCCTGCTG 780  
Db 721 GCTGGGCGAGGATAGAAAGCTGCTGCGCAGGCTCCCTGCCCGGGCCCTGCTG 780  
Qy 781 TCATGGAAGCTGCTGGAGAGCTGCTCCCGCGCGAGCCCTTCCTGCGAGGGTCCACGTCCA 840  
Db 781 TCATGGAAGCTGCTGGAGAGCTGCTCCCGCGCGAGCCCTTCCTGCGAGGGTCCACGTCCA 840  
Qy 841 GGCACCGGGCTGCGACACCCACCCCGCGCGGGGACCTGCGCTGGGTGGCCCTTAAC 900  
Db 841 GGCACCGGGCTGCGACACCCACCCCGCGCGGGGACCTGCGCTGGGTGGCCCTTAAC 900  
Qy 901 CCGGGCGGTAGCTGTTAGATGGCAAGTCTCCGGTCCGGACACCGCAAAACCCCAAT 960  
Db 901 CCGGGCGGTAGCTGTTAGATGGCAAGTCTCCGGTCCGGACACCGCAAAACCCCAAT 960  
Qy 961 CCGGCTGCGCGACCTCTGACCCCGCGCCACCGGACGACAGACTGGGCTCCCGAGC 1020  
Db 961 CCGGCTGCGCGACCTCTGACCCCGCGCCACCGGACGACAGACTGGGCTCCCGAGC 1020  
Qy 1021 CGAGCGGCTGCGGGACACCGGTGCTGGAAACCGAGGACCTTTGTAAACGCGACGTG 1080  
Db 1021 CGAGCGGCTGCGGGACACCGGTGCTGGAAACCGAGGACCTTTGTAAACGCGACGTG 1080  
Qy 1081 TTTGCTCTTTTCAAAAACAAGATAAATGCTTAACTGCTGAAAAGCTTGGCGCCT 1140  
Db 1081 TTTGCTCTTTTCAAAAACAAGATAAATGCTTAACTGCTGAAAAGCTTGGCGCCT 1140  
Qy 1141 AAAAGATGCTGGGTGACTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGACCA 1200  
Db 1141 AAAAGATGCTGGGTGACTAGATGCTAGGATCAGTTTGTGTTTCAATGTAATGGACCA 1200  
Qy 1201 CCGGAGCTCGTACGGACATAGCAGGGGACTGAAGGCTTTCAGGTACTGCTGGTGGG 1260  
Db 1201 CCGGAGCTCGTACGGACATAGCAGGGGACTGAAGGCTTTCAGGTACTGCTGGTGGG 1260  
Qy 1261 GGTGATGCTACAGCGCGATCAGAGTTTGTGCTTCTGGAACCTTGACACTGCACCA 1320  
Db 1261 GGTGATGCTACAGCGCGATCAGAGTTTGTGCTTCTGGAACCTTGACACTGCACCA 1320  
Qy 1321 CGGTAATGCTGAATGACCAATATTTACAGATCACAGCGCATCTTCTTCAACATGA 1380  
Db 1321 CGGTAATGCTGAATGACCAATATTTACAGATCACAGCGCATCTTCTTCAACATGA 1380

Qy 1381 TTTTACACAGTTGACTTAATATGTTGATAATGTAGATCACAATTAACCATACCCAC 1440  
Db 1381 TTTTACACAGTTGACTTAATATGTTGATAATGTAGATCACAATTAACCATACCCAC 1440  
Qy 1441 CTCAGGCTTCTACTTCTGTAATTTTTCAGCAGGTGTTTAACTCTTTGTACCTCAGCTTCT 1500  
Db 1441 CTCAGGCTTCTACTTCTGTAATTTTTCAGCAGGTGTTTAACTCTTTGTACCTCAGCTTCT 1500  
Qy 1501 TCATTACAAAAATAGGGTACTAGCCAGCGGGTGGCTGCGCCTGTATATCCACACT 1560  
Db 1501 TCATTACAAAAATAGGGTACTAGCCAGCGGGTGGCTGCGCCTGTATATCCACACT 1560  
Qy 1561 TGGGAGGCGGAGGAGCCCGGATCACTTGAGGTCAAGATTTTCAGACCGAGCTGGTCAAC 1620  
Db 1561 TGGGAGGCGGAGGAGCCCGGATCACTTGAGGTCAAGATTTTCAGACCGAGCTGGTCAAC 1620  
Qy 1621 ATGGGTGAAACCGCCGCTCTACAAAAATATAAAAACTTTCAGCTGAGTGTGGTAGCCATG 1680  
Db 1621 ATGGGTGAAACCGCCGCTCTACAAAAATATAAAAACTTTCAGCTGAGTGTGGTAGCCATG 1680  
Qy 1681 ACTGTAATCCCAAGCAACTCAGGAGCTGAGGAGAGAAATCGCTTGAACTGGGAGCGGA 1740  
Db 1681 ACTGTAATCCCAAGCAACTCAGGAGCTGAGGAGAGAAATCGCTTGAACTGGGAGCGGA 1740  
Qy 1741 GGTTCGAGTGAAGTGAATCGTACCACTGCACTCCAGCTTGGGCGACAGCGAGCTCT 1800  
Db 1741 GGTTCGAGTGAAGTGAATCGTACCACTGCACTCCAGCTTGGGCGACAGCGAGCTCT 1800  
Qy 1801 GCCTTAAAAATAAATAAATAATTTTAAAAAAATAGGGGTACTAATACTACCTTAAAG 1860  
Db 1801 GCCTTAAAAATAAATAAATAATTTTAAAAAAATAGGGGTACTAATACTACCTTAAAG 1860  
Qy 1861 GATGAGGTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTGTAATCT 1920  
Db 1861 GATGAGGTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCTGTAATCT 1920  
Qy 1921 CAACACTTTGGGAGTCTGTTGGCGGAGGATCACTTGAGCCCGAGGATTTGAGACTAGTCT 1980  
Db 1921 CAACACTTTGGGAGTCTGTTGGCGGAGGATCACTTGAGCCCGAGGATTTGAGACTAGTCT 1980  
Qy 1981 GGGCAACAGACATGCTCTATAGTTGTTGGTTTGTGTTTGTACCAAGGTGTTGGTGT 2040  
Db 1981 GGGCAACAGACATGCTCTATAGTTGTTGGTTTGTGTTTGTACCAAGGTGTTGGTGT 2040  
Qy 2041 GTGCACTGCACTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCGAG 2100  
Db 2041 GTGCACTGCACTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCCGAG 2100  
Qy 2101 GGTGAGGCTGAGTGAGCCATGTTGCGCACTGCACTCCAGCTGGGCAACACAGCAA 2160  
Db 2101 GGTGAGGCTGAGTGAGCCATGTTGCGCACTGCACTCCAGCTGGGCAACACAGCAA 2160  
Qy 2161 GACCTTGTCTCAAAACAAACAAAGCATACTCAATAAGTGTCTGGCTCTCTATATGAT 2220  
Db 2161 GACCTTGTCTCAAAACAAACAAAGCATACTCAATAAGTGTCTGGCTCTCTATATGAT 2220  
Qy 2221 CAATATGCTGGTGGATCTTGAATCTCTGATCTGATCTCATACGATTTTCTGA 2280  
Db 2221 CAATATGCTGGTGGATCTTGAATCTCTGATCTGATCTCATACGATTTTCTGA 2280  
Qy 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAACCGCTCAAGGCACTCCCTCTGCC 2340  
Db 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAACCGCTCAAGGCACTCCCTCTGCC 2340  
Qy 2341 ACCACAAAGCAATTTGATTTTAACTTGAATGTCTTCTTCTTCCAACTTTAGGTAAA 2400  
Db 2341 ACCACAAAGCAATTTGATTTTAACTTGAATGTCTTCTTCTTCCAACTTTAGGTAAA 2400  
Qy 2401 TTAATCTTGGTCAAGGTTCTTGAAACAGCCCTTTTGTGCTATGCTGCAATTTGAATACATGGC 2460  
Db 2401 TTAATCTTGGTCAAGGTTCTTGAAACAGCCCTTTTGTGCTATGCTGCAATTTGAATACATGGC 2460



Db	174214	GGTGTGCGCTACGAGCGGATCAGACAGTTTGTGTCTTTGGAACTTGACACTGACCA	174273
Qy	1321	CGGTAATGCTGAACCTGCAACAATATTAAGATCAGACGCGCATCATCTTCTTCAACATGA	1380
Db	174274	CGGTAATGCTGAACCTGCAACAATATTAAGATCAGACGCGCATCATCTTCTTCAACATGA	174333
Qy	1381	TTTAAACACAGTTGACATTAATATGCTGGATTAATGTTAGATTCACAAATTAACCATACCCAC	1440
Db	174334	TTTAAACACAGTTGACATTAATATGCTGGATTAATGTTAGATTCACAAATTAACCATACCCAC	174393
Qy	1441	CTCAGGCTTCTACTTCTGTAATTTTGACGAGTTGTTTAACTCTTTGTACCTCAGCTTCT	1500
Db	174394	CTCAGGCTTCTACTTCTGTAATTTTGACGAGTTGTTTAACTCTTTGTACCTCAGCTTCT	174453
Qy	1501	TCATTTACAAAATATAGGGTACTAGCAGCGGGGTGCTCGCGCTGTAAATCCAGCACT	1560
Db	174454	TCATTTACAAAATATAGGGTACTAGCAGCGGGGTGCTCGCGCTGTAAATCCAGCACT	174513
Qy	1561	TGGGGAGCGGAGCGGATCACTTGAGGTCAGAGTTTCAGACCCAGCGCTGTTCAAC	1620
Db	174514	TGGGGAGCGGAGCGGATCACTTGAGGTCAGAGTTTCAGACCCAGCGCTGTTCAAC	174573
Qy	1621	ATGGGTGAAACGCGGCTCTACCAAAAATATAAAAACTTAGCTGAGTGGTAGCGCATG	1680
Db	174574	ATGGGTGAAACGCGGCTCTACCAAAAATATAAAAACTTAGCTGAGTGGTAGCGCATG	174633
Qy	1681	ACTGTAATCCAGCAACTCAGGAGGCTGAGCGCAGAGAAATCGCTTGAACCTGGAGCGGA	1740
Db	174634	ACTGTAATCCAGCAACTCAGGAGGCTGAGCGCAGAGAAATCGCTTGAACCTGGAGCGGA	174693
Qy	1741	GGTTGAGTGAGTGATGTTACCACTGCACCTCCAGCTTGGCGGACAGCGAGCTCT	1800
Db	174694	GGTTGAGTGAGTGATGTTACCACTGCACCTCCAGCTTGGCGGACAGCGAGCTCT	174753
Qy	1801	GCCTTAAAAATAAATAAATTTTAAAAAATATAGGGTACTAATATCTACTTTAAAG	1860
Db	174754	GCCTTAAAAATAAATAAATTTTAAAAAATATAGGGTACTAATATCTACTTTAAAG	174813
Qy	1861	GATGAGGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTAATCT	1920
Db	174814	GATGAGGGTTAAATTAAGTACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTAATCT	174873
Qy	1921	CAACACTTTGGAGTCTGTGGCGGAGGATCACTTGAGCCAGGAGTTGAGACTAGTCT	1980
Db	174874	CAACACTTTGGAGTCTGTGGCGGAGGATCACTTGAGCCAGGAGTTGAGACTAGTCT	174933
Qy	1981	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTATACAGGTGTGGTGGT	2040
Db	174934	GGGCAACAGAGACATGCTCTATAGTTGTGTTTGGTTTGTGTTTATACAGGTGTGGTGGT	174993
Qy	2041	GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGTGGGAGACTGCTTACGAGCCAGGA	2100
Db	174994	GTGCACCTGCAGTCCAGCTACTAGGAGGCTGAGTGGGAGACTGCTTACGAGCCAGGA	175053
Qy	2101	GGTGCAGGCTGAGTGAGCCATGATTTGTCACCTGCACCTCCAGCTGGGCAACAGACAA	2160
Db	175054	GGTGCAGGCTGAGTGAGCCATGATTTGTCACCTGCACCTCCAGCTGGGCAACAGACAA	175113
Qy	2161	GACCTTGTCTCAAAAACAAAACAAAGCATACATTAAGTGTCTCGGCTCTATATGATT	2220
Db	175114	GACCTTGTCTCAAAAACAAAACAAAGCATACATTAAGTGTCTCGGCTCTATATGATT	175173
Qy	2221	CAATATGTGGTGGATTCTTGAATCTCTTCTCTGACTCAGACTCTCATAGATTTTCTGA	2280
Db	175174	CAATATGTGGTGGATTCTTGAATCTCTTCTCTGACTCAGACTCTCATAGATTTTCTGA	175233
Qy	2281	ACTTTTGGAGAACTCTTGCCTCTCTGCAATTTGCAAAACCGTCAAAAGCACTCCCTTCGCC	2340
Db	175234	ACTTTTGGAGAACTCTTGCCTCTCTGCAATTTGCAAAACCGTCAAAAGCACTCCCTTCGCC	175293
Qy	2341	ACCACACAAAGCAATTTGATTTTAAACTTGAATCTGTCTCTCTCTTCAAACTTAGGTAA	2400

Db	175294	ACCACACAAAGCAATTTGATTTTAAACTTGACTATGTCTCTTGTTCACAACTTAGGTAA	175353
Qy	2401	TTAATCTTGGTCAGGGTCTCTGAACAGCCCTTTAGTACTACTATGCAATGATACATGCG	2460
Db	175354	TTAATCTTGGTCAGGGTCTCTGAACAGCCCTTTAGTACTACTATGCAATGATACATGCG	175413
Qy	2461	CCTACAGCT 2469	
Db	175414	CCTACAGCT 175422	
RESULT 3			
AC139664/c			
LOCUS	216522 bp	DNA	linear
DEFINITION	Homo sapiens clone RP11-812D13, WORKING DRAFT SEQUENCE, 9 unordered pieces.		HTG 27-MAR-2003
AC139664			
VERSION	AC139664.5	GI:29293970	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 216522)		
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Anaratunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbarta, J., Bencon, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhey, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kravic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louie, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Morgan, M., Morris, S., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Nguyen, A., Nguyen, N., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, K., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 216522)		
AUTHORS	Worley, K.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	3 (bases 1 to 216522)		

# AUTHORS TITLE JOURNAL

Worley, K. C.  
Direct Submission  
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Mar 27, 2003 this sequence version replaced gi:28467047.

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HETI  
Center clone name: RP11-812D13  
----- Summary Statistics

Sequencing vector: Plasmid;  
Chemistry: Dye-terminator; Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 215023 bases at least Q40  
Consensus quality: 215949 bases at least Q30  
Consensus quality: 216649 bases at least Q20  
Estimated insert size: 210467; sum-of-contrigs estimation  
Quality coverage: 10x in Q20 bases; sum-of-contrigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contrigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contrigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1056: contig of 1056 bp in length  
\* 1057 1156: gap of unknown length  
\* 1157 2298: contig of 1142 bp in length  
\* 2299 2398: gap of unknown length  
\* 2399 3437: contig of 1039 bp in length  
\* 3438 3537: gap of unknown length  
\* 3538 5421: contig of 1884 bp in length  
\* 5422 5521: gap of unknown length  
\* 5522 6548: contig of 1027 bp in length  
\* 6549 7859: contig of 1211 bp in length  
\* 7860 7959: gap of unknown length  
\* 7960 21513: contig of 13554 bp in length  
\* 21514 21613: gap of unknown length  
\* 21614 40787: contig of 19174 bp in length  
\* 40788 40888: gap of unknown length  
\* 40889 216522: contig of 175635 bp in length.

## FEATURES source

1. 216522  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-812D13"

## ORIGIN

Query Match 99.1%; Score 2447.4; DB 2; Length 216522;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2462; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
  
Qy 1 TTGAGTCTGTCTGGAGCTCCGGCCAGAGCAGGGCGTATTGTTTCACTCGGTGAATGCT 60  
Db 19780 TTGAGTCTGTCTGGAGCTCCGGCCAGAGCAGGGCGTATTGTTTCACTCGGTGAATGCT 19721  
  
Qy 61 CATTTACGTAAGAAACACGAGCAACGAACTGCCGGAGCGCGGACACCCCGCA 120  
Db 19720 CATTTACGTAAGAAACACGAGCAACGAACTGCCGGAGCGCGGACACCCCGCA 19661  
  
Qy 121 GGGCCGGGTACAGCAGCTGTGTCCAAACAGCGCGGAGGCGCCCGCACCTCCC 180

19660 GGGCCGGGTACAGCAGCTGTGTCCAAACAGCGCGGAGGCGCCCGCCACCTCCC 19601  
181 CCGACCGGGCCGGCCCGGAGCCCTCGCTCGGGGGCTCGAGCGCAACCGGCACACT 240  
19600 CCGACCGGGCCGGCCCGGAGCCCTCGCTCGGGGGCTCGAGCGCAACCGGCACACT 19541  
241 GAGCGAGCGGGCCGGCCCGGAGCGGCTAGCGGGAGGCGCGCGCGGGCGGCTG 300  
19540 GAGCGAGCGGGCCGGCCCGGAGCGGCTAGCGGGAGGCGCGCGCGGGCGGCTG 19481  
301 AGTTACCTGCGCGCGGGAGCTAGAGCGCGGGCGGGCGGGAGCGGGCGCTCGGG 360  
19480 AGTTACCTGCGCGCGGGAGCTAGAGCGCGGGCGGGCGGGAGCGGGCGCTCGGG 19421  
361 CGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAGACTGGGG 420  
19420 CGAACACGCTGGGCACTCCATTGGGGCTGTTTACTCCCAACTCTCGCGAGACTGGGG 19361  
421 GCGGGCCAGAGGCGCCACAGCTCGGAGCTCAGCTCGCGGACCCAGCGTGCCTGTC 480  
19360 ACCGGGCGAGAGGCGCCACAGCTCGGAGCTCAGCTCGCGGACCCAGCGTGCCTGTC 19301  
481 TGTCCCGCGCTCCCGGGGCTTGGTGGCGCTCTGAGCGCGCTGGGCGAGCGGACCG 540  
19300 TGTCCCGCGCTCCCGGGGCTTGGTGGCGCTCTGAGCGCGCTGGGCGAGCGGACCG 19241  
541 CCGGAGGATGGAACGAAGTGTTCGCGACATTTGCGCGCGCGGGCGGGCGGCGAGG 600  
19240 CCGGAGGATGGAACGAAGTGTTCGCGACATTTGCGCGCGCGGGCGGGCGGCGAGG 19182  
601 AAGCGAGGGGCGTGGCCACAGAGCTGCGAGCGCGGAGAACCGCTGGGGGAAACCTTG 660  
19181 AAGCGAGGGGCGTGGCCACAGAGCTGCGAGCGCGGAGAACCGCTGGGGGAAACCTTG 19122  
661 GTCCGCTCTGCGCGTCTTAGATCCCGGAAAGAGCACGGGCGCGGAAAGCGGCCAG 720  
19121 GTCCGCTCTGCGCGTCTTAGATCCCGGAAAGAGCACGGGCGCGGAAAGCGGCCAG 19062  
721 GCTGGCGCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGGCGCTGCTG 780  
19061 GCTGGCGCAGGATCTAGAAAGACTGCTGGCGCAGGCTCCCTGCCCGCGGGCGCTGCTG 19002  
781 TCATGGACTCTGTGGAGAGCTCGCTTCCCGCGCGGAGCCCTTCTGCGAGGGGTCCACGTCA 840  
19001 TCATGGACTCTGTGGAGAGCTCGCTTCCCGCGCGGAGCCCTTCTGCGAGGGGTCCACGTCA 18942  
841 GGCACCGGGGCTCGGACACCCCAACCGCGCGGGCGGACCTGCGCTGGGTGCCCTTAAC 900  
18941 GGCACCGGGGCTCGGACACCCCAACCGCGCGGGCGGACCTGCGCTGGGTGCCCTTAAC 18882  
901 CCGGGCGGTAGCTCGTTAAGATGGGGAAGTGTCCGTCGGGAACACCGCAAAACCCCAAT 960  
18881 CCGGGCGGTAGCTCGTTAAGATGGGGAAGTGTCCGTCGGGAACACCGCAAAACCCCAAT 18822  
961 CCGCGCTGCGCGACCTCTGACCCCGCGGCCCAACCGGAGCAGAGCTGGGCGCTCCGACG 1020  
18821 CCGCGCTGCGCGACCTCTGACCCCGCGGCCCAACCGGAGCAGAGCTGGGCGCTCCGACG 18762  
1021 CGAGCGCGCTGCGGGGACACCGGTGCGTGGAAACCGGAGGACCTTTGTAAACCGCACG 1080  
18761 CGAGCGCGCTGCGGGGACACCGGTGCGTGGAAACCGGAGGACCTTTGTAAACCGCACG 18702  
1081 TTTGCTCTTTTGAATAAACAAGATAAATGTTTAAATGTCTGTAAGAGCTTGGCGCCT 1140  
18701 TTTGCTCTTTTGAATAAACAAGATAAATGTTTAAATGTCTGTAAGAGCTTGGCGCCT 18642  
1141 AAAAGATGCTGGGTGACTTAGATGTAGGATCAGTTTGTGTTTCAATGTAATGACACG 1200  
18641 AAAAGATGCTGGGTGACTTAGATGTAGGATCAGTTTGTGTTTCAATGTAATGACACG 18582  
1201 CCGGAGCTCCGTACGGCACTAGCAGGGGACTGAAAGCGTCTTCAGGTACTGTGGTGGG 1260



```
Db 18581 CCGGACTCCGTACGGCACTAGCAGGGGACTGAAGCGCTTCAGGTACTGCTGGTGGGC 18522
Qy 1261 GGTGATGCGCTACAGGCCGATCAGACAGTTTGTGTCTCTGGAACCTTGACACTGCACCA 1320
Db 18521 GGTGATGCGCTACAGGCCGATCAGACAGTTTGTGTCTCTGGAACCTTGACACTGCACCA 18462
Qy 1321 CGGTAATGCTGAACCTGCAAAATATTTACAGATCAGACAGCCATCATCTTCTTCAACATGA 1380
Db 18461 CGGTAATGCTGAACCTGCAAAATATTTACAGATCAGACAGCCATCATCTTCTTCAACATGA 18402
Qy 1381 TTTAACACAGTTGACTTAATATGGTGGATAAATGTAGAAATCAAAATTAACCATACCCAC 1440
Db 18401 TTTAACACAGTTGACTTAATATGGTGGATAAATGTAGAAATCAAAATTAACCATACCCAC 18342
Qy 1441 CTCAGGCTCTACTTCGTAAATTTTGAAGGTTGTTTAACTTCTTGTACCTCAGCTTCT 1500
Db 18341 CTCAGGCTCTACTTCGTAAATTTTGAAGGTTGTTTAACTTCTTGTACCTCAGCTTCT 18282
Qy 1501 TCATTAACAAAATAGGGGTACTAGCAGGGGGTGGCTCGGCCCTGTAAATCCACGACT 1560
Db 18281 TCATTAACAAAATAGGGGTACTAGCAGGGGGTGGCTCGGCCCTGTAAATCCACGACT 18222
Qy 1561 TGGGAGGCGCGAGCGGACTCACTTGAGGTGAGAGTTTCAGACAGCCCTGTGTCAAC 1620
Db 18221 TGGGAGGCGCGAGCGGACTCACTTGAGGTGAGAGTTTCAGACAGCCCTGTGTCAAC 18162
Qy 1621 ATGGGTGAACCGCGCTCTACAAAATATATAAACTTAGTGTAGTGTGGTAGCGCATG 1680
Db 18161 ATGGGTGAACCGCGCTCTACAAAATATATAAACTTAGTGTAGTGTGGTAGCGCATG 18102
Qy 1681 ACTGTAAATCCAGCACTCAGGAGGCTGAGGAGAGAAATCGCTTGAACCTGGAGGCGGA 1740
Db 18101 ACTGTAAATCCAGCACTCAGGAGGCTGAGGAGAGAAATCGCTTGAACCTGGAGGCGGA 18042
Qy 1741 GGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1800
Db 18041 GGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 17982
Qy 1801 GCCTTAAATATAAATAAATTTTAAATAAATAGGGGTACTAATATCTACCTTAAAG 1860
Db 17981 GCCTTAAATATAAATAAATTTTAAATAAATAGGGGTACTAATATCTACCTTAAAG 17922
Qy 1861 GATGAGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGSCCTGTAACT 1920
Db 17921 GATGAGGTTAAATTAAGTACACATAGCCCTAGCGAGTGGCTTATGSCCTGTAACT 17862
Qy 1921 CAACACTTTGGAGTCTGTGGCGGAGGATCACTTGAGCCAGAGTGTGAGACTAGTCT 1980
Db 17861 CAACACTTTGGAGTCTGTGGCGGAGGATCACTTGAGCCAGAGTGTGAGACTAGTCT 17802
Qy 1981 GGGCAACAGAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTAAACAGGTTGTGTGT 2040
Db 17801 GGGCAACAGAGACATGTCTCTATAGTTGTGTTTGGTTTGTGTTTAAACAGGTTGTGTGT 17742
Qy 2041 GTGCACCTGAGTCCAGCTACTAGGGAGGCTGAGGTGGAGGACTGCCTGAGCCACGGA 2100
Db 17741 GTGCACCTGAGTCCAGCTACTAGGGAGGCTGAGGTGGAGGACTGCCTGAGCCACGGA 17682
Qy 2101 GGTGAGGCTGAGTGAAGCATGATGTGCACTGCACTCCAGCTGGGCAACACAGCAA 2160
Db 17681 GGTGAGGCTGAGTGAAGCATGATGTGCACTGCACTCCAGCTGGGCAACACAGCAA 17622
Qy 2161 GACCTTGTCTCAAAAACAAACAAAGGATCACTATAAGTGTCTGGCTCTCTATATGTT 2220
Db 17621 GACCTTGTCTCAAAAACAAACAAAGGATCACTATAAGTGTCTGGCTCTCTATATGTT 17562
Qy 2221 CAATATGTTGTTGATTTCTTGAATCTTCTGATCTAGATCTCATAGATTTCTGA 2280
Db 17561 CAATATGTTGTTGATTTCTTGAATCTTCTGATCTAGATCTCATAGATTTCTGA 17502
Qy 2281 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAACCGGTCAAAGGCACTCCCTCTGCC 2340
Db 17501 ACTTTTGGAGAACTCTTGGCTCTCTGCAATTTGCAACCGGTCAAAGGCACTCCCTCTGCC 17442
```

```
Qy 2341 ACCACACAAGCATTTGATTTTAACTTGAATATGTTCTTCTGTTCCAACTTTAGTAAA 2400
Db 17441 ACCACACAAGCATTTGATTTTAACTTGAATATGTTCTTCTGTTCCAACTTTAGTAAA 17382
Qy 2401 TTAATCTTGGTCAGGGTCTCTGAACAGCCCTTTAGTCACTATGCAATTGAATACATGCG 2460
Db 17381 TTAATCTTGGTCAGGGTCTCTGAACAGCCCTTTAGTCACTATGCAATTGAATACATGCG 17322
Qy 2461 CCTACAGCT 2469
Db 17321 CCTACAGCT 17313

RESULT 4
AR527527
LOCUS
DEFINITION
Sequence 1 from patent US 6723553.
AR527527
ACCESSION
AR527527.1 GI:53914632
VERSION
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE
1. (bases 1 to 980)
AUTHORS
Abrahamsen, L., Ekblom, J., Forsgren, M., Horling, J., and Johansson, P.
TITLES
Promoter sequences
JOURNAL
Patent: US 6723553-A 1 20-APR-2004;
FEATURES
Location/Qualifiers
1..980
source
/organism="unknown"
/mol_type="genomic DNA"
```

```
Query Match 39.7%; Score 980; DB 6; Length 980;
Best Local Similarity 100.0%; Pred. No. 1e-199;
Matches 980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 353 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 412
Db 1 CTCGCGGGGAACACGCTGGGCACTCCATTCGGGGCTGTTTACTCCCAACTCTCGCGAG 60
Qy 413 ACTGCGGCGCGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 472
Db 61 ACTGCGGCGCGCGGCGAGGCGCCACAGCTGGGAGCCTCAGCTCCGCGGACCCAGCGT 120
Qy 473 GCCCTGTCTGTCCCGCTCCCGGGCTTTCGTCGCGCTCTGAGACGCGTGGGCGAGCGG 532
Db 121 GCCCTGTCTGTCCCGCTCCCGGGCTTTCGTCGCGCTCTGAGACGCGTGGGCGAGCGG 180
Qy 533 GACACGCGCGGAGGATGGAAGAGTGTCTCGGACATTTTGGCGGGCGGGGGCGGGTGG 592
Db 181 GACACGCGCGGAGGATGGAAGAGTGTCTCGGACATTTTGGCGGGCGGGGGCGGGTGG 240
Qy 593 CAGGGTGAAGCGGAGGCGGCTGGCAGCGAGCTGCGAGCGCGCGGAGAACGCGTGGGG 652
Db 241 CAGGGTGAAGCGGAGGCGGCTGGCAGCGAGCTGCGAGCGCGCGGAGAACGCGTGGGG 300
Qy 653 AACCTTGTCTGTCCGCTTTCGCGCTCTAGGATCCCGAAAAGGAGCAAGCGCGCGAAA 712
Db 301 AACCTTGTCTGTCCGCTTTCGCGCTCTAGGATCCCGAAAAGGAGCAAGCGCGCGAAA 360
Qy 713 GCGGCGAGGCTGGGCGGAGGATCTAGAAAGATGCTGGCGGAGGCTCCCTGCCCGCGG 772
Db 361 GCGGCGAGGCTGGGCGGAGGATCTAGAAAGATGCTGGCGGAGGCTCCCTGCCCGCGG 420
Qy 773 GCTGCTGTCTAGGACTCGTGGAGAGCTCGCTTCCGCGGAGACCTTCTTGCAGGGGTC 832
Db 421 GCTGCTGTCTAGGACTCGTGGAGAGCTCGCTTCCGCGGAGACCTTCTTGCAGGGGTC 480
Qy 833 CAGCTCCAGGCAACCGCGGCTCGGACACCCCAACCCCGCGGCGGCACTTCCCTTGGTGC 892
Db 481 CAGCTCCAGGCAACCGCGGCTCGGACACCCCAACCCCGCGGCGGCACTTCCCTTGGTGC 540
```



```
QY 893 CCCTTAACCCGGCGGTAGTCTGTTAAGATGGCGAAGTGTCCGGTCCGGAACACGCGAAA 952
|
|
|
Db 541 CCCTTAACCCGGCGGTAGTCTGTTAAGATGGCGAAGTGTCCGGTCCGGAACACGCGAAA 600
|
|
|
QY 953 CCGCAAAATCCCGCTGCGCGACCTCTGACCCCGCCCGCCCGGACGACGACGACGACGACG 1012
|
|
|
Db 601 CCGCAAAATCCCGCTGCGCGACCTCTGACCCCGCCCGCCCGGACGACGACGACGACGACG 660
|
|
|
QY 1013 TCCCGACGCGACGCGCTGCGGGACACCGGTGCGTGGCAACGAGGACGACGACGACGACG 1072
|
|
|
Db 661 TCCCGACGCGACGCGCTGCGGGACACCGGTGCGTGGCAACGAGGACGACGACGACGACG 720
|
|
|
QY 1073 GCCACGCTGTTGCTCTTTTGAATAAACAAGATAAATGTTTAAACTGTCTGAAAAGCT 1132
|
|
|
Db 721 GCCACGCTGTTGCTCTTTTGAATAAACAAGATAAATGTTTAAACTGTCTGAAAAGCT 780
|
|
|
QY 1133 TGGCGCTAAAGATGCTGGGTGACCTAGATGCTAGGATGATGTTTGGTTTCAATGTAAA 1192
|
|
|
Db 781 TGGCGCTAAAGATGCTGGGTGACCTAGATGCTAGGATGATGTTTGGTTTCAATGTAAA 840
|
|
|
QY 1193 TGGACACGCGGACCTCGGTACGCGACTAGCAGGAGCTGAAAGCGCTTTCAGGTACTGC 1252
|
|
|
Db 841 TGGACACGCGGACCTCGGTACGCGACTAGCAGGAGCTGAAAGCGCTTTCAGGTACTGC 900
|
|
|
QY 1253 TGGTGGCGGTGATGCGCTACAGCGCGATCAGACAGTGTGTTGCTCTTCTGGAATTGACA 1312
|
|
|
Db 901 TGGTGGCGGTGATGCGCTACAGCGCGATCAGACAGTGTGTTGCTCTTCTGGAATTGACA 960
|
|
|
QY 1313 CTGCACACCGGTAACTGCTGA 1332
|
|
|
Db 961 CTGCACACCGGTAACTGCTGA 980
|
|
|
RESULT 5
BC052334 532 bp mRNA linear PRI 05-MAY-2003
LOCUS Homo sapiens cDNA clone IMAGE:6176570, partial cds.
DEFINITION BC052334
ACCESSION BC052334
VERSION BC052334.1 GI:30354085
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 532);
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,W.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Ketterman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalios,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
MEDLINE
PUBMED
2 (bases 1 to 532)
Strausberg,R.
AUTHORS
Direct Submission
TITLE
Submitted (02-MAY-2003) National Institutes of Health, Mammalian
```

REMARK  
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland:  
Web site: <http://www.nisec.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Aklter,N., Ayle,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,  
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,  
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,  
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,  
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,  
Tsurgon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,  
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAC Plate: 115 Row: n Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis, Genomescan gene prediction.

FEATURES  
source

1..532  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6176570"  
/tissue\_type="uterus, leiomyosarcoma"  
/clone\_lib="NIH\_MGC\_71"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 21.1%; Score 521.4; DB 9; Length 532;  
Best Local Similarity 99.8%; Pred. No. 2.6e-101;  
Matches 522; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 624 GCTGCCAGCGCGGAGAACGCGCTGGGGAAACCTTGGTCCGCTCTGCGCTCGCTCTAG 683  
Db 1 GCTGCCAGCGCGGAGAACGCGCTGGGGAAACCTTGGTCCGCTCTGCGCTCGCTCTAG 60  
QY 684 GATCCCCGAAAAAGGAGACACGCGCGGCGGAAAGCGGCGGAGGATCTAGAAAGAC 743  
Db 61 GATCCCCGAAAAAGGAGACACGCGCGGCGGAAAGCGGCGGAGGATCTAGAAAGAC 120  
QY 744 TGCTGGCGGAGGCTCCCTGCCCCCGGGGCTGTGTGTATGAGCTCTGGAGAGTCCG 803  
Db 121 TGCTGGCGGAGGCTCCCTGCCCCCGGGGCTGTGTGTATGAGCTCTGGAGAGTCCG 180  
QY 804 TTCCCGCGGGGACCTTCTCTGAGGCGGTCCAGTCCAGGCACCGCGGCTCGGACACCCC 863  
Db 181 TTCCCGCGGGGACCTTCTCTGAGGCGGTCCAGTCCAGGCACCGCGGCTCGGACACCCC 240  
QY 864 ACCCCCGCGCGGCGACCTGCGCTGGGTGCCCTTAAACCCGGGCGGTAGCTCGTTAAGATG 923  
Db 241 ACCCCCGCGCGGCGACCTGCGCTGGGTGCCCTTAAACCCGGGCGGTAGCTCGTTAAGATG 300  
QY 924 GCGAAGTGTCCGGTCCGGAACACCGGAAACCCCAATCCCGCTCCCGGACCTCTGACC 983  
Db 301 GCGAAGTGTCCGGTCCGGAACACCGGAAACCCCAATCCCGCTCCCGGACCTCTGACC 360  
QY 984 CCGGCGCCCGGAGGACGAGACTGGGCTCCGAGCGCGAGCGGCTGCGGGGACACCG 1043  
Db 361 CCGGCGCCCGGAGGACGAGACTGGGCTCCGAGCGCGAGCGGCTGCGGGGACACCG 420

QY 1044 GTGCGTGGCAACGAGGACCTTTGTAACGCCACGTTGCTCTTTTGAACAAACAAAG 1103  
 Db 421 GTGCGTGGCAACGAGGACCTTTGTAACGCCACGTTGCTCTTTTGAACAAACAAAG 480  
 QY 1104 AATAAATGTTAAATGCTGCTGAAAGCTTGGCGCTTAAAGA 1146  
 Db 481 AATAAATGTTAAATGCTGCTGAAAGCTTGGCGCTTAAAGA 523

RESULT 6  
 HS358H7/c  
 LOCUS  
 DEFINITION Human DNA sequence from clone Rp3-358H7 on chromosome X, complete sequence.  
 ACCESSION 277249  
 VERSION 277249.2 GI:26800831  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 JOURNAL 1 (bases 1 to 148864)  
 Pavitt,R.

Direct Submission  
 Submitted (05-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Dec 14, 2002 this sequence version replaced gi:1460066.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute

Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
 Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX  
 Rp3-358H7 is from the library RPCI-3 constructed by the group of  
 Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pCVPAC2.

## FEATURES

source  
 1. .148864  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="RZPD:RPCIP704H07358"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="Rp3-358H7"  
 /clone\_lib="RPCI-3"

## ORIGIN

Query Match 14.1%; Score 347; DB 9; Length 148864;

Best Local Similarity 70.7%; Pred. No. 1.2e-63;  
 Matches 503; Conservative 0; Mismatches 205; Indels 3; Gaps 3;

QY 1509 AAAATAGGGGTAATAAGCAGCGGGGGTCTCGCCCTGTGTAATCCAGACACTTTGGGAGG 1568  
 Db 77689 AAAATGTTAGGGGTGGGCCAGGTGTGGTCTCACACCTGTGTAATCCCAACACTTTGGGAGG 77630  
 QY 1569 CCGAGGACCGGATCACTTGGAGTGCAGAGTTTCAGACCGCCTGGTCAACATGGGTGA 1628  
 Db 77629 CCGAGGCGGGTGGATCACTTGGAGTGTGGGACCAAGCTGTCCAACAT-GGTGA 77571  
 QY 1629 AACGCGGGCTTACCAAAAATAATAAACTTAGCTAGTGTGGTAGCGCATGACTGTGTAAT 1688  
 Db 77570 AACCCGGTCTTACTTAAAAACACAAAAATTAGCCAGGGTGCAGCAGGTGCCAGGAAT 77511  
 QY 1689 CCCAGCAACTCAGGAGGTGAGGCA-GAGAAATCGCTTGAACCTGGGAGCGGAGGTGCA 1747  
 Db 77510 CCCAGCTACTCGGAGGGCTGAGCGAGGAGAAATCGCTTGAACCCGGGAGACAGAGGTGCA 77451  
 QY 1748 GTGAGCTGAGATCGTACCACCTGCATCCAGCTTGGGCGAC-AGAGCGAGACTCTGCCTTA 1806  
 Db 77450 GTGAGCCAAGATCGTCCCACTGCATCCAGCTTGGGTGACGAGAGGAAATCCCGTCTCA 77391  
 QY 1807 AAAATAAATAAATAAATTTTAAAAAAAATAGGGGTACTAATATCTACCTTAAAGGATGAG 1866  
 Db 77390 AAAAAAAAATAAATAAATAAAGTATGGGTGATGTCTTCAAGTGGCATCTACCCGTGGAAT 77331  
 QY 1867 GGTAAATTAAGTACACATTAAGCCCTAGCCAGTGGCTTATGCTGTAAATCTCAACAC 1926  
 Db 77330 TGCCTCTTTAGGAGGATGGTAGAGGCCAGGCACAGTGGCTCACACCTGTCATCCCAACAC 77271  
 QY 1927 TTTGGAGTCTGTGGGGGAGGATCACTTGGAGCCAGGAGTTTGAGACTAGTCTGGGCAA 1986  
 Db 77270 TTTGGGATGCTGGGGTGGGAGGATTCCTTGGCCAGGAAATTCAGAGACTAGCTTGGGCAA 77211  
 QY 1987 CAGAGACATGCTCTATAGTGTGTTGGTTGTTGTTTATCCAGAGTGTGGTGTGCAC 2046  
 Db 77210 GATAGTGACACCCCTCTATACAAATAGTTAAATAACAGTCAGTCATGGCGTGCACAC 77151  
 QY 2047 CTGCACTCCAGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCAGAGAGTGA 2106  
 Db 77150 CTGTAGTCTCTGCTACTTGGGAGGCTGAGGTGGGAGGATCACATGATGTCATGAGTTGA 77091  
 QY 2107 GCCTCAGTGGCCATGATTGTCACCTGCATCTCCAGCTTGGGCAACAGCAAGACCTT 2166  
 Db 77090 GGTCTCAGTGGGCTGATGCTGCTCCTGCTCCTATATG 2217  
 QY 2167 GTCTCAAAACAAAACAAAAGCATACTCATAAAGTGTCTCGGCTCTCTATATG 2217  
 Db 77030 GTCTCAAAACAAAACAAAAGAAAAGAAAAGAGGAGGATGGTAAAGG 76980

RESULT 7  
 AC010519  
 LOCUS Homo sapiens chromosome 19 clone LLNL-261D9, complete sequence.  
 DEFINITION AC010519  
 ACCESSION AC010519  
 VERSION AC010519.8 GI:21747432  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 46741)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 46741)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 46741)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 46741)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 5 (bases 1 to 46741)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jul 14, 2002 this sequence version replaced gi:16195205.  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.8% of Sequence;  
 Estimated Total Number of Errors is 0.1.

FEATURES  
 source  
 1. .46741  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="19"  
 /clone="L1NLR-261D9"

ORIGIN  
 Query Match 14.0%; Score 346; DB 9; Length 46741;  
 Best Local Similarity 71.4%; Pred. No. 1.7e-63;  
 Matches 510; Conservative 0; Mismatches 200; Indels 4; Gaps 4;

QY 1480 CCTCTTTGTACCTCAGCTTCTTCATTACAAAATAGGGTACTAGCCAGCGGGTGGCT 1539  
 DB 40713 CTTCTTTGTAGCCTCATCTGCAAAAATGTGGATCAATCACCTTGGGCGAGCGCGTGGCT 40772

QY 1540 CGGCGCTGTATCCAGCACTTGGGAGCGCGGAGCGGATCACTTGGGCTCAGAG 1599  
 DB 40773 CACACCTGTATCCAGCGCTTTGGGAGCGCAAGCGAGGTAGTCACTTGGGCGCAGGAG 40832

QY 1600 TTTTCAGACAGCTGGTCAACATGGGTGAAACCGCGCTCTACAAAATATAAAACATT 1659  
 DB 40833 TTGGAGACAGCTGGCCACAT-GGTGAACCCCGCTCTACAAAACACAAAATTT 40891

QY 1660 AGCTGAGTGTGGTAGCGCATCTGTAATCCAGCACTCAGAGGCTGAGGAG-AGAA 1718  
 DB 40892 AGCTGGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 1778

QY 1719 TCGTGTGAACCTGGAGCGCGAGGTTCAGTGAGCTGAGATCGTACCCTGCACTCCAGC 1778  
 DB 40952 TTGCTTAAACCTGGAGCGGAGGTTCAGTAAAGCGAGACCGTGGCCACTGCATTCCAGC 41011

QY 1779 TTGGCGCAGACAGCGAGCTCTGCTTTAAAAATAAATAAATAATTTTAAAAAAATAGG 1838  
 DB 41012 CTGGGCAACAGAGCGAGCTCTGCTTTAAAAATAAATAAATAAATAAAGGATCAGATCA 41071

QY 1839 GGTACTAATATCTACCTTAAAGGATGAGGGTTAAATTAAGTACACACATAAGCCCTAGCG 1898  
 DB 41072 CCATGGTCTATGCCATGGTGTGCTGAGGACTTGGAAATAGCAC-CTGGTGGCGGGCG 41130

QY 1899 CAGTGGCTTATGCTGTAACTCAACACTTTGGAGTCTGTGGCGGAGATCACTTGAG 1958  
 DB 41131 CAGTGTCTCAGCGCTGTAAATCCAGCACCTTTGGGAGGCTTGGGCGGCGAGATCACTTGAG 41190

QY 1959 CCGAGGAGTTGAGCATGTCTGGGCAACAGACATGTCTCTATAGTTGTGTTGTTT 2018  
 DB 41191 GCCAGAGTTGAGACAGACCGCTGGCCCAACATGGTGAACCCCGTCTCTACTAAATAACA 41250

QY 2019 TGTATTTACAGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2078  
 DB 41251 AAAATTAGCTGGGTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 41310

QY 2079 GGAGGACTCCTGAGCCCGAGGAGTTCGAGCTGCGAGTGCAGTGCATGATTTGTCCTGCAC 2138  
 DB 41311 GGAGAAATCATTGAATCTCGGAGGCGGAGGTTGCGAGTGCAGTGCATGATTTGTCCTGCAC 41370

QY 2139 TCCAGCCTGGGCAAC-ACAGCAAGACCTTGTCTCAAAACAAACAAACAAACGATA 2191  
 DB 41371 TCCAGCCTGGGCGCAACAGCAAAACTCTGTCTCAAAACAAACAAACAAACAAAAA 41424

RESULT 8  
 AC093324/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome RPCI-11 clone RP11-565M3, WORKING DRAFT  
 SEQUENCE, 43 unordered pieces.  
 AC093324  
 AC093324.1 GI:15209236  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 194241)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 194241)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT  
 ----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H\_NH0565M03  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 0%  
 Sequencing vector: plasmid; 100%  
 Chemistry: Dye-primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 171387 bases at least Q40  
 Consensus quality: 176465 bases at least Q30  
 Consensus quality: 179780 bases at least Q20  
 Insert size: 210000; agarose-fp  
 Insert coverage: 190041; sum-of-contigs  
 Quality coverage: 3.09 in Q20 bases; agarose-fp  
 Quality coverage: 3.56 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 43 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1215: contig of 1215 bp in length  
 \* 1216: gap of unknown length  
 \* 1316: contig of 1810 bp in length  
 \* 3225: gap of unknown length  
 \* 4721: contig of 1496 bp in length  
 \* 4821: gap of unknown length  
 \* 4822: contig of 1235 bp in length  
 \* 6057: gap of unknown length  
 \* 6157: contig of 1743 bp in length  
 \* 7900: gap of unknown length  
 \* 8000: contig of 1392 bp in length





Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L24580  
 Center clone name: 2116\_G\_1  
 -----

FEATURES	Location/Qualifiers
source	1..167587
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="15"
	/map="15"
	/clone="CTD-2116G1"
	/clone_lib="CITDI Human BAC"
	complement(11..144)
repeat_region	/rpt_family="L2"
repeat_region	complement(145..356)
repeat_region	/rpt_family="AluSg/x"
repeat_region	complement(357..1109)
repeat_region	/rpt_family="L2"
repeat_region	3713..3828
repeat_region	/rpt_family="L1ME"
repeat_region	complement(3834..4006)
repeat_region	/rpt_family="MIR"
repeat_region	4574..4603
repeat_region	/rpt_family="(GAAA)n"
repeat_region	complement(4976..5167)
repeat_region	/rpt_family="MIR"
repeat_region	5639..5775
repeat_region	/rpt_family="L1"
repeat_region	6745..6766
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(8290..8370)
repeat_region	/rpt_family="MIR3"
repeat_region	9077..9288
repeat_region	/rpt_family="MIR"
repeat_region	9496..10838
repeat_region	/rpt_family="L1MD1"
repeat_region	10839..11162
repeat_region	/rpt_family="AluSx"
repeat_region	11163..12229
repeat_region	/rpt_family="L1MD1"
repeat_region	complement(13118..13414)
repeat_region	/rpt_family="AluY"
repeat_region	13667..14061
repeat_region	/rpt_family="L2"
repeat_region	14366..14395
repeat_region	/rpt_family="AT_rich"
repeat_region	14748..15052
repeat_region	/rpt_family="AluSx"
repeat_region	complement(15603..15768)
repeat_region	/rpt_family="MIR"
repeat_region	complement(15878..15932)
repeat_region	/rpt_family="L2"
repeat_region	complement(15988..16069)
repeat_region	/rpt_family="MER81"
repeat_region	complement(16309..16442)
repeat_region	/rpt_family="MER46A"
repeat_region	complement(16557..16730)
repeat_region	/rpt_family="MIR"
repeat_region	complement(16746..16830)
repeat_region	/rpt_family="MIR"
repeat_region	16845..16939
repeat_region	/rpt_family="MER117"
repeat_region	complement(17267..17436)
repeat_region	/rpt_family="MIR3"
repeat_region	complement(17914..18014)
repeat_region	/rpt_family="MER103"
repeat_region	complement(19124..19415)
repeat_region	/rpt_family="AluJo"
repeat_region	complement(19571..19636)

repeat_region	/rpt_family="L2"	19637..19926
repeat_region	/rpt_family="AluJo"	complement(19927..19984)
repeat_region	/rpt_family="L2"	complement(20449..21788)
repeat_region	/rpt_family="L2"	21985..22144
repeat_region	/rpt_family="MIR"	22728..22838
repeat_region	/rpt_family="Charlie8"	23014..23044
repeat_region	/rpt_family="AT_rich"	23140..23256
repeat_region	/rpt_family="FLAM_C"	23265..23350
repeat_region	/rpt_family="(TA)n"	23357..23381
repeat_region	/rpt_family="AT_rich"	24908..25206
repeat_region	/rpt_family="AluSx"	complement(25805..25979)
repeat_region	/rpt_family="L1MB4A"	complement(26043..26262)
repeat_region	/rpt_family="L1MB4A"	complement(26861..26993)
repeat_region	/rpt_family="FLAM_C"	complement(27023..27317)
repeat_region	/rpt_family="AluSx"	27952..28300

Query Match 13.9%; Score 343.4; DB 9; Length 167587;  
 Best Local Similarity 75.3%; Pred. No. 6.9e-63;  
 Matches 507; Conservative 0; Mismatches 156; Indels 10; Gaps 6;

QY	1519	TACTAGCCAGGGGGTGGCTCGCGCTGTAAATCCAGACACTTGGGGAGGCGGAGCAGC	1578
DB	83398	TACTGGCCAGGCGCAGTGGCTCAGCGCTGTAAATCCAGACACTTGTGAGGCGGAGGTGG	83457
QY	1579	CGATCAGCTTGAGTTCAGAAGTTTCAGACCAGCTGGTCAACATGGGTGAACCGCGCT	1638
DB	83458	TGGATCAGCTTGAGTTCAGAGGTTTCAGACCAGCTGGTCAACATGGGTGAACCGCGCT	83516
QY	1639	CTACCAAAAATATAAACTTAGCTGAGTGTGAGCGCATGACTTAATCCAGCAACT	1698
DB	83517	CTACTAAAAATACAAAAA-TTAGCTGGGTGTGGTGGGGTGAAGTAAATCCAGCAACT	83575
QY	1699	CAGGAGGCTGAGGCA-GAGATCGCTTGAACCTGGGAGGCGGAGGTTCAGTGAAGTGA	1757
DB	83576	CAGGAGGCTGAGGAGGAGATCACTTGAACCGGAGGCGGAGGTTCAGTGAAGTGAAG	83635
QY	1758	ATCGTACCACTGACCTCCAGCTTGGGCGACAGAGGAGACTCTGCTTAAAAATAAATAA	1817
DB	83636	ATTGCACCACTGCACTCCAGCTGGGTGACAGAGTGAGACTCCGTCTCAAAAACAATAA	83695
QY	1818	ATAATTTTAAAAAATAAGGGTACTAATATCTACTTTAAAGGATGAGGGTTAAATTA	1877
DB	83696	ATAATAATAAATAAATAAGAGCTTCC-CTTAAAGGCAAGGTCAAGAGATATGAAGA	83753
QY	1878	GTACACACATAAGCCCTAGCGAGTGGCTTATGCCTGTAAATCTCAACACTTTGGGAGTCT	1937
DB	83754	TTCAAGATGGAGGCGGCGCACAGTGGCTTATGCCTGTAAATACCAATTTCTTTGGGAGGC	83813
QY	1938	GTGGCGGAGGATCACTTGAGCCAGGAGTTGAGACTAGTCTGGGCAACAGAGACATGT	1997
DB	83814	AAGGTGGGTGATCACCTGAGGTGAGAGTTTATAGACAGGCTGGGCAACATGGTGAAC	83873
QY	1998	CTCTATAGTTGTTTGGTTTGTGTTTAA-CAGAGTGTGGTGTGTGTCACCTGCAGTCCC	2056
DB	83874	CCCTCTCTACTAAAAATAATAAATAATAGCTGGGTGGTGGTGGGATGCTGTATATCC	83933
QY	2057	AGCTACTAGGAGGCTGAGGTGGGAGGACTGCTGAGCCAGGAGGTGAGGCTGAGGTG	2116



```

Db      83934 AGCTACTCAGGAGCTTGGGAGGAGAAATTGCTTGAACCCGGAGGTGGAGCTGCAGTG 83993
QY      2117 AGCCATTGTTGCCACTGCACCTCCAGCTGGGCAACA----CAGCAACAGCTTGTCTCA 2172
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      83994 AGCCAGATTGAGCTGCTGCACCTCCAGCTGGGCAACAGAGCAAAAACCTGTCTCA 84053
QY      2173 AAACAAACAAA 2185
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      84054 AAAAAAAAAAAAA 84066

RESULT 10
AC074125/c
LOCUS   AC074125      52479 bp      DNA      linear      HTG 15-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone CTB-53M18, WORKING DRAFT SEQUENCE,
15 unordered pieces.
ACCESSION AC074125
VERSION   AC074125.1 GI:9211235
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 52479)
          Doe Joint Genome Institute.
          Direct Submission
          Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
          -----Genome Center
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov
          -----
Project Information
Center Project Name: 84862, BC49885
Center clone name: CIT978SKB_53M18
          -----
Summary Statistics
Consensus quality: 39219 bases at least Q40
Consensus quality: 46049 bases at least Q30
Consensus quality: 48400 bases at least Q20
Estimated insert size: 121910; agarose-fp estimation
Quality coverage: 5.54 in Q20 bases; agarose-fp estimation
Quality coverage: 13.21 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1269: contig of 1269 bp in length
* 1270: gap of unknown length
* 1369: contig of 1620 bp in length
* 2989: gap of unknown length
* 3089: contig of 1040 bp in length
* 4123: gap of unknown length
* 4223: contig of 1666 bp in length
* 5895: gap of unknown length
* 5995: contig of 2300 bp in length
* 8295: gap of unknown length
* 8395: contig of 1730 bp in length
* 8396: gap of unknown length
* 10125: contig of 1616 bp in length
* 10225: gap of unknown length
* 11841: contig of 1616 bp in length
* 11941: gap of unknown length
* 13973: contig of 2032 bp in length
* 14073: gap of unknown length

```

```

* 14074 16722: contig of 2649 bp in length
* 16723 16822: gap of unknown length
* 16823 20125: contig of 3303 bp in length
* 20126 20225: gap of unknown length
* 20226 26887: contig of 6662 bp in length
* 26887 26987: gap of unknown length
* 26988 28929: contig of 1942 bp in length
* 28930 29029: gap of unknown length
* 29030 37888: contig of 8859 bp in length
* 37889 45112: contig of 7124 bp in length
* 45113 52479: contig of 7267 bp in length.
* 52479: contig of 7267 bp in length.
FEATURES
      source
      1..52479
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="19"
      /clone="CTB-53M18"
      /clone_lib="CalTech human BAC library B"

ORIGIN
Query Match      13.7%; Score 339.4; DB 2; Length 52479;
Best Local Similarity 72.6%; Pred. No. 4.5e-62;
Matches 496; Conservative 0; Mismatches 176; Indels 11; Gaps 4;

QY      1519 TACTAGCCAGGCGGTGGCTCGCCCTGTATATCCAGCACCTTGGGGAGCCGAGGCAGC 1578
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31328 TCAGCGCGGCGATATGGCTCAGCCCTGTATATCCAGAACCTTTGGAGGCTGAGGCAGG 31269
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1579 CGGATCCTTGAGGTGAGAAAGTTTCAGACCGCTGGTCAACATGGGTGAACCGCGGCT 1638
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31268 CGGATCCTTGAGGTGAGAGGTGAGAGCCGCTGGCCAAAAT-GGGGAAACTCCGTCT 31210
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1639 CTACCAAAAATATAAAAACCTTAGCTGAGTGTGGTAGCGGATCAGCTGTATATCCAGCAACT 1698
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31209 CTACTGAAAAATACAAAAA-TTAGCCGCGCATGGTGGCGCATGCCGTGTATCCAGTACT 31151
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1699 CAGGAGGCTGAGGCA-GAGAAATCGTTGAACTCTGGGAGCGGAGGTTCAGTGCAGCTGAG 1757
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31150 CCGGAAGCTGAGGCGATGAGAAATCACTTCAACAGGAGGAGGAGGTTGCACTGAGCTGAG 31091
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1758 ATCTGACCACTGCACCTCCAGCTTGGGCGACAGAGGAGACTCTGCTTAAAAATAAATAA 1817
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31090 ACCTCACCACCTGCACCTCCAGCCTAGGCGACAGAGTGAGAAAGCTGTCTTAAATAAATAA 31031
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1818 ATAAATTTTAAAAAAATAGGGGTACTATATCTTACCTTAAAGGATGAGGTTAAATTTAA 1877
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      31030 TAAATAAAAACCTCTTTTGACCCCTCTTAGAATCTCTCCACAAATGCAGAAAAGAAATTA 30971
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1878 GTACACACATAAGCCCTAGCGAGTGCTTATGCTGTATCTCAACACTTTTGGGAGTCT 1937
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      30970 -----ATGGGTCTGGCATGGGGCTCACACCTGTATATCCAGCACTTTGGGAGGCC 30919
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1938 GTGGCGGAGGATCACTTGAGCCCGAGGAGTTTGAAGTCTGTGGGCAACAGAGACATGT 1997
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      30918 GAGGACAGGAGATCTCTTGAGCCCGAGGAGTTTGAGACCGCGCTGGCAACACAGTGAAC 30859
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1998 CTCCTATAGTTGTTTGGTTTGTGTTTATACAGGTGTGTTGTGTCACCTGCAGTCCCA 2057
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      30858 CCGCTCTCACAAAAAAATTTAAAAAGTAGCTGCAATAGTGTGTCACCTGTAGTCTCA 30799
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      2058 GCTACTAGGAGGCTGAGGTGGAGGAGCTGCTGAGCCCGAGGAGTTCGAGGCTGCAGTGA 2117
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      30798 GCTACTTGGAGGCTGAGGTGGAGGAGTGGCTTGGAGCTGGGAGGTTGAGGCTGCAGTGA 30739
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      2118 GCCATGATGTGCACCTGCACCTCCAGCTGGGCAACACAGCAAGACCTTGTCTCAAAAAC 2177
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      30738 GCTGTGATCATGCCACTGCACCTTCACCTGGGCTACACATCGAGACCTGTCTCAAAAAA 30679
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      2178 AAACAAAAAGCATACTCATAAAG 2200
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



Db 30678 AAAAAAAAAAAAAAAAAAGAG 30656

RESULT 11  
AC010605  
LOCUS  
DEFINITION Homo sapiens chromosome 19 clone CTB-189B5, complete sequence.  
ACCESSION AC010605  
VERSION AC010605.4 GI:7019286  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL  
REFERENCE 1 (bases 1 to 86684)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
JOURNAL  
REFERENCE 2 (bases 1 to 86684)  
TITLE DOE Joint Genome Institute.  
AUTHORS Direct Submission  
JOURNAL  
REFERENCE 3 (bases 1 to 86684)  
TITLE Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint  
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
REFERENCE 4 (bases 1 to 86684)  
TITLE Submitted (22-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell  
AUTHORS Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
REFERENCE 5 (bases 1 to 86684)  
TITLE Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell  
AUTHORS Drive, Walnut Creek, CA 94598, USA  
JOURNAL  
COMMENT On Feb 22, 2000 this sequence version replaced gi:6910541.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.7% of Sequence;  
Estimated Total Number of Errors is 0.5.

FEATURES  
source  
1. .86684  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="CTB-189B5"

ORIGIN  
Query Match 13.7%; Score 339.4; DB 9; Length 86684;  
Best Local Similarity 72.6%; Pred. No. 4.7e-52;  
Matches 496; Conservative 0; Mismatches 176; Indels 11; Gaps 4;

QY 1519 TACTAGCAGCGGGTGGCTCGCGCTGTAATCCAGCAGCTTGGGGAGGCCGAGCAGC 1578  
Db 38703 TCCAGCGCGGCATGATGGCTCAGCGCTGTAATCCAGAGACTTGGGAGCGTGGCAGG 38762  
QY 1579 CGATACATTGAGGTGAGAAGTTTCAGACGAGCTGTCAACATGGGTGAACCGCGCT 1638  
Db 38763 CGGATCACTTGAGTCAAGAGTGTGAGACGAGCTGTGCCAAAAT-GGCGAAATCCCGTCT 38821  
QY 1639 CTACCAAAATATATAAACTTACTGATGTGTGTAGCGCATGCTGTAATCCCGCAACT 1698  
Db 38822 CTACTGAAAATACAAAAA-TTAGCCGGGCATGTGTGGCGCATGCTGTGATCCCGAGTACT 38880  
QY 1699 CAGGAGCTGAGCA-GAGAAATCGCTTGAACCTGGAGGGCGGAGGTTGCAAGTGAAGTGA 1757  
Db 38881 CGGAGAGCTGAGCATGAGATCACTTCAACAGAGGGGCGAGAGTTGCAAGTGAAGTGA 38940  
QY 1758 ATCGTACCACTGCACTCCAGCTTGGGGGAGAGAGAGATCTGCGCTTAAAAATAATAA 1817  
Db 38941 ACCTCACCACGTCACTCCAGCTTAGGCGAGAGAGTGAAGAGCTGTCTTAAATAATAAAA 39000

QY 1818 ATAATTTTTAAAAAATAGGGGTACTATATATCTACCTTAAAGGATGAGGGTTAAATTTAA 1877  
Db 39001 TAAATAAAACCTCTTTTGACCCCTCTTAGAATCTCTCCACAAATGCAGAAAGATTAA 39060  
QY 1878 GTACACACATAAGCCCTAGCGCAGTGGCTTATGCGCTGTGAATCTCAACATTTGGGAGTCT 1937  
Db 39061 -----ATGGGTCTCTGGCATGGGGGCTTCACACCTGTAAATCCAGCACTTTGGGAGGCC 39112  
QY 1938 GTGGCGGAGGATCACTTGGAGCCAGGAGTTTGAGACTAGTCTGGGCAACAGACATGT 1997  
Db 39113 GAGCGAGGAGGATCTCTTGGAGCCAGGAGTTTGAGACCAAGCGCTGGCAACAGTGAAC 39172  
QY 1998 CTCATATAGTTGTGTTGTTGTTTATACCAAGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2057  
Db 39173 CCGTCTCTCAAAAAAATTTAAAAAGTAGCTGCAATAGTGTGTGTGTGTGTGTGTGTGTGT 39232  
QY 2058 GCTACTAGGAGGCTGAGGTGGAGAGCTGCTGAGGCCAGAGGTCGAGGTCGAGTGA 2117  
Db 39233 GCTACTTGGAGGCTGAGGTGGAGAGTGGCTTGGAGCTGGAGGTTGAGGCTGAGTGA 39292  
QY 2118 GCATGATGTGCTCACTGCACTCCAGCTGGGCAACAGAGAGAGCTGTCTCAAAAAC 2177  
Db 39293 GCTGTGATCATGCCACTGCACTTCAAGCTGGGCTGAGTGTGAGAGCCCTGTCTCAAAAAA 39352  
QY 2178 AAACAAAAAGCATACTCATAAAG 2200  
Db 39353 AAAAAAAGAAAAAAGAAAAAG 39375

RESULT 12  
AC080005  
LOCUS  
DEFINITION Homo sapiens 3 BAC RP11-475N22 (Roswell Park Cancer Institute Human  
ACCESSION AC080005  
VERSION AC080005.24 GI:22024355  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alshrocks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbata,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., He,X.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kurehi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulsegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,  
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,  
Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,  
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,  
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,



/function="low quality"									
repeat_region	13623..13639	/rpt_family=" (ATGGTG)n"							
repeat_region	13640..14059	/rpt_family="MSTA"							
repeat_region	14060..14120	/rpt_family=" (ATGGTG)n"							
repeat_region	14127..14543	/rpt_family="L2"							
repeat_region	complement(14803..15276)	/rpt_family="MLTIP"							
repeat_region	15281..15439								
Query Match 13.7%; Score 339.4; DB 9; Length 169583;									
Best Local Similarity 59.6%; Pred. No. 5e-62;									
Matches 517; Conservative 0; Mismatches 221; Indels 5; Gaps 4;									
QY	1447	CTTCTACTTCGTAATTTGAGCAGGTTGTTTAACCTCTTTGTACCTCAGCTCAGCTTCTTCATTA	1506						
DB	163181	CCTCAATTCCTTTTTTAAGAGTTATTTCTGTGCCGAGCAGTGCCCTCTCTCTTTC	163240						
QY	1507	CAAAATAGGGTACTAGCAGCGGGGTGGCTGGCCCTGTAAATCCAGACACTTGGGA	1566						
DB	163241	AAAAATCAGAAAGCTGCTGGGTGCTGGCTCACACCTGTAACTCCAGCACTTTAGGA	163300						
QY	1567	GGCCGAGGCGGATCACTTGAGCTCAGAGTTTCAGACCAGCCTGCTCAACATGGCT	1626						
DB	163301	GGCCGAGGCGGAGATCACTTGAGCTCAGAGTTTCAGACCAGCCTGCTCAACATGGCT	163359						
QY	1627	GAACCCCGGCTTACCAAAATATATAAACTTAGCTGAGTGGTGGTGGCATGACTGTA	1686						
DB	163360	AAACCTCATCTTTACTTAAATAACAAAATAGCAGATGCTGCTGCTCTATA	163419						
QY	1687	ATCCAGCAACTCAGAGGCTGAGGA-GAGAACTGCTGAACTGGGAGGCGAGGTTG	1745						
DB	163420	ATCCAGCTACTCGGAGGCTGAGGAGGAGAAATGCTTGAACCCAGGAGGAGAGTTTG	163479						
QY	1746	CAGTGAGCTGAGATCTTACACTGCTCCAGCTTGGGGGACAGGCGAGCTCTGCCCTT	1805						
DB	163480	CAGTGAGCAGAGTTGTGCACCTGCACTCAGCTGGCCGACAGAGCGAGA--CTGTCTG	163537						
QY	1806	AAAAATAATAATAATTTTAAAAAATAGGGTACTAATATCTACCTTAAAGGATGA	1865						
DB	163538	AAAAAATAATAATAATAGAAAGCTGAAGTCTTTGTGCCCCCAAGTCATAGAGACGAA	163597						
QY	1866	GGTTAAATTAAGTACACATAGCCCTAGCCAGTGGCTTATGCTGTATATCTCAACA	1925						
DB	163598	AGAAAAACGAAAAAGAAAGCTGGGGGGGACAGTGGCTCATGCTCTGTAAATCCCAACA	163657						
QY	1926	CTTTGGGAGTCTGGCGGAGGATCACTTGAGCCAGGAGTTTGAGACTAGCTCTGGGA	1985						
DB	163658	CTTTGGGAGGCAAGTGGGAGATACCTTGAGCCAGGAGTTGAGAGACAGGCTGGCCA	163717						
QY	1986	ACAGAGACATGCTCTATAGTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTGCA	2045						
DB	163718	GCATAG-CAAAACCTGCTCTACAAAAATACAAAAATAGCAGTGTGGTGTGTA	163776						
QY	2046	CTTGAGTCCAGCTACTAGGAGGCTGAGTGGAGGACTGCTCCTGAGCCCGAGAGGTCG	2105						
DB	163777	CCTGTGCTCTAGCTACTTGGGATGCTGAGTGGGAGGATGCTTAAGCCTGGGAGGTGG	163836						
QY	2106	AGCTTCAGTGAGCCATGATTGGCCACTGCACTCCAGCTGGGACACAGCAAGACCT	2165						
DB	163837	AGGTGCGAGGAGCCAGATTTGGCCACTGCACTCCAGCTGGGTCACAGTGGAGACCT	163896						
QY	2166	TGCTCAAAAAACAAAAAGC	2188						
DB	163897	TGCTCAAAAAAATAAATGC	163919						
RESULT 13									
AC020715/c									
LOCUS									
	AC020715	197992 bp	DNA	linear	HTG	17-AUG-2000			

DEFINITION	Homo sapiens chromosome 19 clone RP11-446K10, WORKING DRAFT
ACCESSION	AC020715
VERSION	AC020715.4 GI:9838201
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Waterston,R.H.
TITLE	The sequence of Homo sapiens clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 197992)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (08-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Aug 17, 2000 this sequence version replaced gi:8570194.
----- Genome Center -----	
Center: Washington University Genome Sequencing Center	
Center code: WUGSC	
Web site: http://genome.wustl.edu/gsc/index.shtml	
----- Project Information -----	
Center project name: H_NH0446K10	
----- Summary Statistics -----	
Sequencing vector: M13; #	
Sequencing vector: plasmid; #	
Chemistry: Dye-primer ET; # of reads	
Assembly: Dye-terminator Big Dye; # of reads	
Assembly program: Phrap; version 0.990319	
Consensus quality: 193859 bases at least Q40	
Consensus quality: 194948 bases at least Q30	
Consensus quality: 195923 bases at least Q20	
Insert size: 209000; agarose-fp	
Insert size: 197594; sum-of-contigs	
Quality coverage: 6.86 in Q20 bases; agarose-fp	
Quality coverage: 7.28 in Q20 bases; sum-of-contigs	
-----	
* NOTE: This is a 'working draft' sequence. It currently	
* consists of 7 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
* 1 1582: contig of 1582 bp in length	
* 1583 1582: gap of unknown length	
* 1683 10849: contig of 9167 bp in length	
* 10850 10849: gap of unknown length	
* 10950 26324: contig of 15375 bp in length	
* 26325 26424: gap of unknown length	
* 26425 49772: contig of 23348 bp in length	
* 49773 49872: gap of unknown length	
* 49873 78964: contig of 29092 bp in length	
* 78965 79064: gap of unknown length	
* 79065 107990: contig of 28926 bp in length	
* 107991 108090: gap of unknown length	
* 108091 197992: contig of 89902 bp in length.	
FEATURES	
source	
1..197992	
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="19"	
/clone="RP11-446K10"	
1..1582	
misc_feature	
/note="assembly_name:Contig4"	
1683..10849	
misc_feature	
/note="assembly_name:Contig6"	



MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 19, 2002 this sequence version replaced gi:17647019.  
All repeats were identified using RepeatMasker:  
Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L9215

Center clone name: 635\_N\_19

#### FEATURES

##### source

1. .173252

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-635N19"

/clone\_lib="RPC1-11 Human Male BAC"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region 7124..7442  
/rpt\_family="MER65A-int"  
repeat\_region 7443..7892  
/rpt\_family="MER65A"  
repeat\_region 8797..8975  
/rpt\_family="MER5A"  
repeat\_region 9249..9351  
/rpt\_family="MIR"  
repeat\_region 9719..9880  
/rpt\_family="LIMB1"  
complement(10160..10263)  
/rpt\_family="MIR3"  
repeat\_region 10453..10523  
/rpt\_family="MER94"  
complement(10639..10937)  
/rpt\_family="AluSx"  
complement(10809..10983)  
/note="single clone coverage"  
complement(10841..10854)  
/note="<30 qual single clone coverage"  
complement(10889..10901)  
/note="<30 qual single clone coverage"  
complement(10913..10933)  
/note="<30 qual single clone coverage"  
complement(10970..10978)  
/note="<30 qual single clone coverage"  
complement(10972..11226)  
/rpt\_family="AluSx"  
repeat\_region 11234..11275  
/rpt\_family="AT-rich"  
repeat\_region 11529..12097  
/rpt\_family="MER114"  
repeat\_region 12300..12578  
/rpt\_family="LIMASA"  
repeat\_region 15024..15055  
/rpt\_family="(T)n"  
repeat\_region 15336..15723  
/rpt\_family="MER109"  
repeat\_region 16785..16865  
/rpt\_family="LJ"  
repeat\_region 17662..17716  
/rpt\_family="(ATG)n"  
complement(17753..17944)  
/rpt\_family="MER91A"  
repeat\_region 18683..18982  
/rpt\_family="AluY"  
complement(19042..19306)  
/rpt\_family="AluSx"  
complement(19307..19393)  
/rpt\_family="MER103"  
complement(20344..20790)  
/rpt\_family="LIMAA4"

Query Match 13.7%; Score 338.4; DB 9; Length 173252;

Best Local Similarity 71.7%; Pred. No. 8.2e-62;

Mismatches 499; Conservative 0; Mismatches 191; Indels 6; Gaps 4;

QY 1508 AAAATAGGGGTACTAGCCAGCGGGTGGCTCGCGCCCTGTATCCAGCATTGGGGAG 1567  
DB 170208 AGAACTAAGTATATTGGCCAGGCATGGTGGCTCACGCCCATATATCACAGGACTTTGGGAG 170149  
QY 1568 GCGAGGACCGGATCATTGAGGTCAGAGTTTCAGACCCAGCCCTGGTCAACATGGGTG 1627  
DB 170148 GCCAGGCGAGGAGATCATTGAGGTCAGAGTTTGGGCTGACCCCTGACACAAACAT -GGTA 170090  
QY 1628 AAACCGCGCTCTACCAAAAAT--ATAAAAATTAGCTGAGTGTGTAGCGCATGACTGT 1685  
DB 170089 AAACCTGTCTCTACTATAAAATACAAAAAATAGCCGGTGTGTGGCACACACTGT 170030  
QY 1686 AATCCAGCAACTCAGGAGGCTGAGGCA -GAGATCGTCTGAACCTGGGAGCGGAGGTT 1744  
DB 170029 AATCCAGCTACTCAGGAGGCTGAGGAGAGAAATCGTTGAACCCGAGAGGTGGAGTT 169970





```
repeat_region complement(10651..10682)
repeat_region /rpt_family="MER33"
repeat_region 10683..10986
repeat_region /rpt_family="AluY"
repeat_region complement(10987..11250)
repeat_region /rpt_family="MER33"
repeat_region 11308..11375
repeat_region /rpt_family="L2"
repeat_region 12718..12891
repeat_region /rpt_family="MIR"
repeat_region complement(12999..13079)
repeat_region /rpt_family="L2"
repeat_region 13088..13738
repeat_region /rpt_family="L2"
repeat_region 13441
repeat_region /note="Probably a C"
repeat_region complement(14105..14154)
repeat_region /rpt_family="L2"
repeat_region 14160..14189
repeat_region /rpt_family="(A)n"
repeat_region 14218..14246
repeat_region /rpt_family="(TG)n"
repeat_region 14767..15154
repeat_region /rpt_family="LMB3"
repeat_region 15792..16076
repeat_region /rpt_family="AluSg"
repeat_region 17327..17519
repeat_region /rpt_family="MIR"
repeat_region 17901..18088
repeat_region /rpt_family="L2"
repeat_region 18207..18227
repeat_region /rpt_family="AT rich"
repeat_region complement(18341..18479)
repeat_region /rpt_family="LIPAL3"
repeat_region 19251..19304
repeat_region /rpt_family="MIR"
repeat_region 19541..19867
repeat_region /rpt_family="AluJo"
repeat_region 20031..20144
repeat_region /rpt_family="L2"
repeat_region 20197..20415
repeat_region /rpt_family="L2"
repeat_region complement(20507..20711)
repeat_region /rpt_family="MER30"
repeat_region 22505..22536
repeat_region /rpt_family="(CA)n"
repeat_region complement(22609..22669)
repeat_region /rpt_family="L2"
repeat_region 23043..23297
repeat_region /rpt_family="L2"
repeat_region 23470..23598
repeat_region /rpt_family="MER2"
repeat_region complement(23600..23915)
repeat_region /rpt_family="AluSx"
repeat_region 23924..23987
repeat_region /rpt_family="MER2"
repeat_region 23935..23941
repeat_region /note="<30 qual SINGL region"
repeat_region 23974..24230
repeat_region /note="<30 qual SINGL region"
repeat_region 24223
repeat_region /note="Probably a G"
repeat_region 24242..24425
repeat_region /rpt_family="L2"
repeat_region 27566..27586
repeat_region /rpt_family="AT rich"
repeat_region complement(28883..28973)
repeat_region /rpt_family="MER5A"
repeat_region 32167..32197
repeat_region /rpt_family="GC rich"
repeat_region 32284..32313
repeat_region /rpt_family="GC rich"
repeat_region 33790..33817

/rpt_family="GC rich"
33827..33878
/rpt_family="(CGG)n"
complement(33975..34047)
/note="<30 qual SINGL region"
34104..34228
/rpt_family="GC rich"
34442..34486
/rpt_family="(TG)n"
complement(35364..35660)
/rpt_family="AluSg"
complement(35766..35887)
/rpt_family="FLAM_A"
37108..37419
/rpt_family="AluJb"
37637..37946

Query Match 13.7%; Score 338.4; DB 9; Length 188923;
Best Local Similarity 71.7%; Pred. No. 8.3e-62;
Matches 499; Conservative 0; Mismatches 191; Indels 6; Gaps 4;

QY 1508 AAAAATAGGGTACTAGCCAGCGGGGTGGCTCGCGCCTGTATATCCAGCACTTGGGGAG 1567
Db 72247 AGAACTAAGTATATTGGCCAGGCATGGTGGCTCAGCCCATATATCACAGGACTTTGGGAG 72306
QY 1568 GCGAGGCGCGGATCACTTGAGGTCAGAACTTTCAGACCAGCCTGGTCAACATGGGTG 1627
Db 72307 GCCAAGGCGAGGAGATCACTGAGGTGAGGATTTGAGACTAGCTGACAAACAT-GGTA 72365
QY 1628 AAACGCGGCTCTACCAAAAT--ATAAACTTATAGTGTGGTGTAGCGCATGACTGT 1685
Db 72366 AAACCTGTCTCTACTATAAAATACAAAAAATTAGCGGGTGTGGTGGCACACACTGT 72425
QY 1686 ATCCAGCACTCAGGAGGCTGAGGCA-GAAGATCGTTTGAACCTTGGAGCGGAGGTT 1744
Db 72426 AATCCAGCTACTCAGGAGGCTGAGCGAGGAAATCGTTGAACCCGAGAGGTGGAGTT 72485
QY 1745 GCAGTGAGTGTGATCGTACCACTGCCTCCAGCTTGGCGACAGCAGCACTCTGCT 1804
Db 72486 GCAGTGAGCAAGATGGCCACCTGCCTCCAGCTTGGGTGACAGAGAGACTCTGTCT 72545
QY 1805 TAAAAATAATAATAATTTTAAAAAATAGGGGTACTAATATCTACTTAAAGGATG 1864
Db 72546 CAAAAAATAAAAAAAGAACTAAACATA--TATTATTATTAGAGTTATGTGCAATA 72603
QY 1865 AGGTTAAATTAAGTACACACATAAGCCCTAGCGAGTGGCTTATGCCCTGTATCTCAAC 1924
Db 72604 TATGGTAAATTTAAAAAGAAAAAGAGCGAGGCAAGTGGCTCACTCTATAATACAGC 72663
QY 1925 ACTTTGGGAGTCTGTGGCGGAGGATCACTTGCAGCCAGGAGTTTGGAGACTAGTCTGGGC 1984
Db 72664 ACTTTGGGAGGAGGTTAGCGGAGTGTCTTGAAGCCAGGAGTTGAAGACCCAGCTGGGC 72723
QY 1985 AACAGAGACATGTCTCTATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGC 2044
Db 72724 AACATGGCGAAACCCCATCTCTACCAAAATAACAAAAATAGCCATGATGGACGT 72783
QY 2045 ACCTGAGTCCAGCTACTAGGAGGCTGAGGTGGAGAGTGCCTTGAGCCAGGAGGTC 2104
Db 72784 GTCTATAATCTCAGCTACTAGGAGGCTGAGGTGGAGAGTGTCTTGAAGCCAGGAGGTC 72843
QY 2105 GAGGCTGAGTGGAGCATGATGTGCCACTGCCAGCTCCAGCTGGGCAACACAGCAGACCC 2164
Db 72844 GAGGCTGAGTGGAGTGTGATCGACCTCCAGCTGGATGACAGATTTGAGACA 72903
QY 2165 TTGTCTCAAAAAACAAAAAGACATCTCAATAAG 2200
Db 72904 CTGCCTCAAAAAAATAAAAAAAGGAAAG 72939

Search completed: August 12, 2005, 02:41:45
Job time : 10570 secs
```